UNIVERSIDADE ESTADUAL PAULISTA

FACULDADE DE CIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ANTONIO CARLOS TEIXEIRA BLAIA JUNIOR

Geração Automática de *Trimap* para *Matting* Digital de Imagens e Vídeo

Bauru
2013

ANTONIO CARLOS TEIXEIRA BLAIA JUNIOR

Geração Automática de *Trimap* para *Matting* Digital de Imagens e Vídeo

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, Área de Concentração - Sistemas de Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de São José do Rio Preto.

Orientador: Prof. Dr.Antonio Carlos Sementille

Bauru
2013

ANTONIO CARLOS TEIXEIRA BLAIA JUNIOR

Geração Automática de *Trimap* para *Matting* Digital de Imagens e Vídeo

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, Área de Concentração - Sistemas de Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de São José do Rio Preto.

Banca Examinadora

Prof. Dr. Antonio Carlos Sementille
UNESP - Bauru
Orientador

Prof. Dr. João Fernando Marar
UNESP - Bauru

Prof. Dr. Ildeberto Aparecido Rodello
FEARP/USP - Ribeirão Preto

Bauru
04 de Fevereiro de 2013

Agradecimentos

Primeiramente a Deus,

Por ter permitido que eu começasse o mestrado e dado força e persistência para eu terminar. Também por tudo que Ele tem me concedido. Sem Ele na minha vida, nada seria possível.

Aos meus familiares e minha noiva,

Pelo constante apoio, incentivo, pelo amor que sempre deu força para continuar e pela compreensão nas horas de desespero e nas horas de mau humor, quando algo não dava certo.

Ao meu orientador, professor Dr. Antonio Carlos Sementille,

Por ter acreditado em mim, e ser uma pessoa muito ponderada e humana. Estando presente em um dos momentos mais díficeis de minha vida me orientando tanto academicamente e conselhos para minha vida pessoal.

Aos meus colegas Henrique e Daniel,

Por compartilhar os momentos difíceis, as mudanças, as viagens. Também pela amizade, lealdade e companheirismo.

A UNESP de Bauru,

Por proporcionar tanto a minha graduação em ciência da computação quanto o mestrado.

Enfim agradeço a todos aqueles que estiveram presentes em algum momento de minha caminhada e contribuíram para o meu crescimento.

Resumo

Blaia Junior, Antonio Carlos Teixeira. **Geração Automática de *Trimap* para *Matting* Digital de Imagens e Vídeo**. Bauru, 2013. 125p. Dissertação de Mestrado. Departamento de Computação, Faculdade de Ciências, Universidade Estadual Paulista.

Técnicas de composição de imagens, com extração de objetos em primeiro plano para serem combinados com um novo plano de fundo, são muito utilizadas em aplicações que vão desde simples montagens em fotos até produções cinematográficas. Estas técnicas são chamadas de *matting* digital. Com elas é possível minimizar custos das produções, uma vez que o ator não precisa estar realmente no local onde a cena irá acontecer. Esta característica também viabiliza sua utilização em programas realizados para a televisão digital, que demandam uma alta qualidade da imagem.

Diversas abordagens de *matting* digital utilizam marcações feitas nas imagens (como informações extras), para determinar o que é o primeiro plano, o plano de fundo e as áreas de incerteza. O conjunto destas informações extras é chamado de *trimap*, definida por um mapa triplo representando a segmentação da imagem de entrada. O *trimap* é feito, geralmente, a partir de marcações manuais realizadas pelo usuário.

Levando em consideração este contexto, o presente trabalho objetivou o desenvolvimento de dois métodos para a geração de *trimaps* de maneira automática (sem interação do usuário), baseando-se, para isto, em informações de profundidade e/ou cores dos *pixels* das imagens de entrada. Os métodos desenvolvidos foram implementados e comparados com suas versões manuais, tendo apresentado resultados quantitativamente semelhantes, para imagens. Os métodos revelaram, também, potencial para serem aplicados em vídeos.

Palavras–chave

Matting, *Chromakey*, *Trimap*

Abstract

Blaia Junior, Antonio Carlos Teixeira. . Bauru, 2013. 125p. Dissertação de Mestrado. Departamento de Computação, Faculdade de Ciências, Universidade Estadual Paulista.

Techniques of composing images with extracting objects in the foreground to be combined with a new background, are widely used in applications ranging from simple assemblies on photos to filmmaking. these techniques are called digital matting. With them you can minimize costs of production, since the actor does not need to be really where the scene will happen. this feature also enables its use in programs made for television digital, which demand a high quality image.

Several approaches utilize digital matting marks made on images (such as extra information), to determine what is the foreground, the background and areas of uncertainty. All these extra information is called a trimap, defined for a triple map representing the segmentation of the input image. The trimap is made, generally, from markings made by the user manuals.

Considering this context, this study aimed to develop Two methods for generating trimaps either automatically (without interaction user), relying, for this, in depth information and / or colors of the pixels the input images. The developed methods were implemented and compared with their manual versions and tested quantitatively similar for images. The methods also revealed the potential to be applied in videos.

Keywords

Matting, Chromakey, Trimap

Sumário

Lista de Figuras **9**

Lista de Tabelas **17**

Lista de Abreviaturas **18**

1 Introdução **19**

1.1 Contextualização 19

1.2 Objetivos do Trabalho 21

1.3 Organização do Trabalho 21

2 *Matting* Digital **22**

2.1 Considerações Iniciais 22

2.2 Definições e Conceitos Iniciais 22

2.3 Os problemas do *matting* digital 26

2.4 *Matting* com informação extra 27

2.4.1 *Chromakey* 28

2.4.2 *Flash matting* 31

2.4.3 *Matting* com Vetor de Câmeras 32

2.4.4 Câmeras *time of flight* 34

2.5 Considerações Finais 36

3 Trimap **37**

3.1 Considerações Iniciais 37

3.1.1 *Stroke Trimap* 38

3.1.2 *Filled Trimap* 39

3.2 *Trimap* automático . . . 39
3.3 Métodos de *Matting* digital . . . 39
3.3.1 *Bayesian Matting* . . . 39
3.3.2 *Robust Matting* . . . 41
3.3.3 *Geodesic Matting* . . . 42
3.3.4 *Geodesic* Modificado . . . 42
3.4 Considerações Finais . . . 43

4 *Kinect* 44
4.1 Considerações Iniciais . . . 44
4.2 Componentes internos do *Kinect* . . . 44
4.3 Mapa de profundidade . . . 45
4.4 *Real-Time Human Pose Recognition* . . . 46
4.5 Pacote de Desenvolvimento de Software para o *Kinect* . . . 46
4.6 Considerações Finais . . . 47

5 Métodos de Automatização de *Trimap* para imagens e vídeo 48
5.1 Considerações Iniciais . . . 48
5.2 Método automático de *trimap* baseado em mapa de profundidade . . . 49
5.3 Método automático de *trimap* baseado em *chromakey* . . . 50
5.4 Considerações Finais . . . 51

6 Implementação dos métodos propostos 52
6.1 Considerações Iniciais . . . 52
6.2 Implementação do método de *trimap* baseado em mapa de profundidade e *chromakey* . . . 52
6.3 Implementação do método de *trimap* baseado em *chromakey* . . . 56
6.4 Considerações finais . . . 58

7 Experimentos e Análise de Resultados 59
7.1 Considerações Iniciais . . . 59
7.2 Métricas utilizadas . . . 61
7.2.1 *Strutural Similarity* - SSIM . . . 61
7.2.2 *Universal Image Quality Index* - UIQI . . . 62

7.3 Resultados utilizando o *trimap* gerado a partir do mapa de profundidade . 62
7.3.1 Análise da qualidade do *trimap* automático 71
7.4 Resultados utilizando o *trimap* gerado por *chromakey* 71
7.4.1 Análise da qualidade do *trimap* automático 77
7.5 Testes em vídeo 78
7.6 Considerações finais 79

8 Conclusões e Trabalhos Futuros **80**
8.1 Conclusões 80
8.2 Trabalhos Futuros 81

Referências Bibliográficas **82**

A Imagens dos testes **85**

Lista de Figuras

Figura 2.1 Composição óptica. Fonte: (MONACO, 2000) 24

Figura 2.2 Exemplo de *matting* com o uso da tela azul de fundo, onde (a) é o objeto à ser extraído, (b) objeto extraído e (c) composição final. Fonte: (SMITH; BLINN, 1996) 25

(a) 25

(b) 25

(c) 25

Figura 2.3 Composição de nove objetos. Fonte: (SMITH; BLINN, 1996) . . . 26

Figura 2.4 Exemplo de objeto com superfície irregular e com sombra. Fonte: (MONTABONE, 2010) 27

Figura 2.5 Exemplo de *matting* considerando a sombra do objeto em primeiro plano, onde (a) imagem entrada, (b) plano de fundo, (c) composição final preservando a sombra do objeto. Fonte: (CHUANG, 2004) 27

(a) 27

(b) 27

(c) 27

Figura 2.6 Exemplo de *chromakey*, onde (a) imagem original, (b) composição final. Fonte: (FOSTER, 2010) 28

(a) 28

(b) 28

Figura 2.7 Cubo RGB com a intersecção do plano S com a pirâmide de inclinação OBTPQ. Fonte: (BERGH; LALIOTI, 1999). 29

Figura 2.8 Exemplo do problema do color spill, onde (a) imagem original, (b) composição final. (MCGUIRE; MATUSIK; YERAZUNIS, 2006) 30

(a) Imagem de entrada 30

(b) Composição final 30

Figura 2.9 Exemplo do Flash Matting. (SUN et al., 2006) 31

(a) Imagem com flash . . . 31
(b) Imagem sem flash) . . . 31
(c) Matte . . . 31
(d) Composição final) . . . 31
Figura 2.10 Exemplo do flash *matting* para um objeto complexo. Fonte: (SUN et al., 2006) . . . 32
Figura 2.11 Vetor com 8 camêras e sistema em tempo real. Matte gerado com um quarto da resolução VGA. Fonte: (JOSHI; MATUSIK; AVIDAN, 2006) 33
Figura 2.12 Vídeo *matting* natural. (JOSHI; MATUSIK; AVIDAN, 2006) . . . 34
(a) Imagem de entrada . . . 34
(b) imagem com a variação das refocadas sintéticas da imagem (região mais escura onde houve menor variação) . . . 34
(c) *Trimap* gerado a partir da variação da imagem . . . 34
(d) Matte . . . 34
(e) Composição final . . . 34
Figura 2.13 Uso da câmera TOF para a extrair o primeiro plano do plano de fundo. (WANG et al., 2010) . . . 35
(a) Câmera ZCam . . . 35
(b) Imagem de profundidade . . . 35
(c) Imagem colorida . . . 35
(d) Resultado da segmentação . . . 35
Figura 3.1 Exemplo do trimap. Fonte: (RHEMANN et al., 2008) . . . 37
(a) Imagem de entrada com as marcações realizadas pelo usuário . . . 37
(b) Trimap, sendo a região B o plano de fundo, região F é o primeiro plano e a região U é a região desconhecida (região indeterminada) . 37
Figura 3.2 Exemplo do trimap. Fonte: (HU, 2010) . . . 38
(a) Imagem de entrada . . . 38
(b) Marcações do usuário) . . . 38
(c) Trimap . . . 38
Figura 3.3 Análise local do pixel desconhecido. Fonte: (CHUANG et al., 2001) 40
Figura 3.4 Exemplo dos resultados do *Bayesian* matting. Fonte: (CHUANG et al., 2001) . . . 41
Figura 3.5 Análise do pixel desconhecido. Fonte: (WANG; COHEN, 2007b) . 42

Figura 3.6 Resultados obtidos com o Geodesic Matting. Fonte: (BAI; SAPIRO, 2009) 42

Figura 4.1 Imagem da parte interna do *Kinect* Fonte: (KUHN, 2011) 45

Figura 4.2 Exemplo das imagens capturadas pelo *Kinect* 46

(a) Mapa de profundidade 46

(b) Imagem da câmera RGB 46

Figura 5.1 Cubo RGB com a interseção dos planos S1 e S2 (ARRUDA, 2012). 50

Figura 6.1 Pipeline do funcionamento da geração do *trimap* a partir do mapa de profundidade. 53

Figura 6.2 Trecho de código relativo à inicialização do *Kinect*. 53

Figura 6.3 Trecho de código que ilustra a leitura dos quadros da câmera RGB e mapa de profundidade do *Kinect*. 54

Figura 6.4 Trecho de código que verifica se o *pixel* pertence a um ser humano. 54

Figura 6.5 Trecho de código responsável pela calibração dos *pixels* das imagens RGB e mapa de profundidade. 55

Figura 6.6 Trecho de código que ilustra a implementação do *chromakey* com base no método de (BERGH; LALIOTI, 1999). 57

Figura 6.7 Trecho de código que ilustra a implementação do *chromakey* modificado para geração do *trimap* automático. 58

Figura 7.1 Ambiente experimental. 60

(a) 60

(b) 60

Figura 7.2 Imagem de entrada capturada pela câmera RGB do *Kinect* com fundo verde. 63

Figura 7.3 Mapa de profundidade gerado pelo *Kinect*. 63

Figura 7.4 *Trimap* gerado pelo método desenvolvido. 64

Figura 7.5 Recorte gerado pelo *Bayesian matting* (CHUANG et al., 2001), com o *trimap* automático. 65

Figura 7.6 Recorte gerado pelo *Robust matting* (WANG; COHEN, 2007b), com o *trimap* automático. 65

Figura 7.7 Recorte gerado pelo Geodesic modificado (ARRUDA, 2012), com o *trimap* automático. 66

Figura 7.8 *Trimap* gerado de forma manual. 66

Figura 7.9 Recorte gerado pelo *Bayesian matting* (CHUANG et al., 2001), com o *trimap* manual. 67

Figura 7.10 Recorte gerado pelo *Robust matting* (WANG; COHEN, 2007b), com o *trimap* manual. 67

Figura 7.11 Recorte gerado pelo Geodesic modificado (ARRUDA, 2012), com o *trimap* manual. 68

Figura 7.12 Composição gerado pelo *Bayesian matting* (CHUANG et al., 2001), com o novo plano de fundo. 68

Figura 7.13 Composição gerado pelo *Robust matting* (WANG; COHEN, 2007b), com o novo plano de fundo. 69

Figura 7.14 Composição gerado pelo Geodesic modificado (ARRUDA, 2012), com o novo plano de fundo. 69

Figura 7.15 Câmera fullHD acoplada ao *Kinect*. 72

Figura 7.16 Imagem capturada pela câmera fullHD. 72

Figura 7.17 *Trimap* automático gerado a partir da imagem RGB da câmera fullHD. 73

Figura 7.18 Recorte gerado pelo *Bayesian matting* (CHUANG et al., 2001), com o *trimap* automático. 73

Figura 7.19 Recorte gerado pelo *Robust matting* (WANG; COHEN, 2007b), com o *trimap* automático. 74

Figura 7.20 Recorte gerado pelo Geodesic modificado (ARRUDA, 2012), com o *trimap* manual. 74

Figura 7.21 Composição gerado pelo *Bayesian matting* (CHUANG et al., 2001), com o novo plano de fundo. 75

Figura 7.22 Composição gerado pelo *Robust matting* (WANG; COHEN, 2007b), com o novo plano de fundo. 75

Figura 7.23 Composição gerado pelo Geodesic modificado (ARRUDA, 2012), com o novo plano de fundo. 76

Figura 7.24 Representação de um dos quadros do vídeo após o *matting* digital com *trimap* automático. 79

Figura A.1 Img_1 - Imagem capturada pela câmera RGB 85

Figura A.2 Img_1 - Trimap manual 85

Figura A.3 Img_1 - Trimap automático 86

Figura A.4 Img_1 - Composição obtida com o Bayesian Matting a partir do Trimap manual 86

Figura A.5 Img_1 - Composição obtida com o Bayesian Matting a partir do Trimap automático 87

Figura A.6 Img_1 - Composição obtida com o Robust Matting a partir do Trimap manual . . . 87
Figura A.7 Img_1 - Composição obtida com o Robust Matting a partir do Trimap automático . . . 88
Figura A.8 Img_1 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual . . . 88
Figura A.9 Img_1 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático . . . 89
Figura A.10 Img_2 - Imagem capturada pela câmera RGB . . . 89
Figura A.11 Img_2 - Trimap manual . . . 90
Figura A.12 Img_2 - Trimap automático . . . 90
Figura A.13 Img_2 - Composição obtida com o Bayesian Matting a partir do Trimap manual . . . 91
Figura A.14 Img_2 - Composição obtida com o Bayesian Matting a partir do Trimap automático . . . 91
Figura A.15 Img_2 - Composição obtida com o Robust Matting a partir do Trimap manual . . . 92
Figura A.16 Img_2 - Composição obtida com o Robust Matting a partir do Trimap automático . . . 92
Figura A.17 Img_2 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual . . . 93
Figura A.18 Img_2 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático . . . 93
Figura A.19 Img_3 - Imagem capturada pela câmera RGB . . . 94
Figura A.20 Img_3 - Trimap manual . . . 94
Figura A.21 Img_3 - Trimap automático . . . 95
Figura A.22 Img_3 - Composição obtida com o Bayesian Matting a partir do Trimap manual . . . 95
Figura A.23 Img_3 - Composição obtida com o Bayesian Matting a partir do Trimap automático . . . 96
Figura A.24 Img_3 - Composição obtida com o Robust Matting a partir do Trimap manual . . . 96
Figura A.25 Img_3 - Composição obtida com o Robust Matting a partir do Trimap automático . . . 97
Figura A.26 Img_3 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual . . . 97

Figura A.27 Img_3 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático 98
Figura A.28 Img_4 - Imagem capturada pela câmera RGB 98
Figura A.29 Img_4 - Trimap manual 99
Figura A.30 Img_4 - Trimap automático 99
Figura A.31 Img_4 - Composição obtida com o Bayesian Matting a partir do Trimap manual 100
Figura A.32 Img_4 - Composição obtida com o Bayesian Matting a partir do Trimap automático 100
Figura A.33 Img_4 - Composição obtida com o Robust Matting a partir do Trimap manual 101
Figura A.34 Img_4 - Composição obtida com o Robust Matting a partir do Trimap automático 101
Figura A.35 Img_4 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual 102
Figura A.36 Img_4 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático 102
Figura A.37 Img_5 - Imagem capturada pela câmera RGB 103
Figura A.38 Img_5 - Trimap manual 103
Figura A.39 Img_5 - Trimap automático 104
Figura A.40 Img_5 - Composição obtida com o Bayesian Matting a partir do Trimap manual 104
Figura A.41 Img_5 - Composição obtida com o Bayesian Matting a partir do Trimap automático 105
Figura A.42 Img_5 - Composição obtida com o Robust Matting a partir do Trimap manual 105
Figura A.43 Img_5 - Composição obtida com o Robust Matting a partir do Trimap automático 106
Figura A.44 Img_5 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual 106
Figura A.45 Img_5 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático 107
Figura A.46 Img_6 - Imagem capturada pela câmera RGB 107
Figura A.47 Img_6 - Trimap manual 108
Figura A.48 Img_6 - Trimap automático 108

Figura A.49 Img_6 - Composição obtida com o Bayesian Matting a partir do Trimap manual . . . 109
Figura A.50 Img_6 - Composição obtida com o Bayesian Matting a partir do Trimap automático . . . 109
Figura A.51 Img_6 - Composição obtida com o Robust Matting a partir do Trimap manual . . . 110
Figura A.52 Img_6 - Composição obtida com o Robust Matting a partir do Trimap automático . . . 110
Figura A.53 Img_6 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual . . . 111
Figura A.54 Img_6 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático . . . 111
Figura A.55 Img_7 - Imagem capturada pela câmera RGB . . . 112
Figura A.56 Img_7 - Trimap manual . . . 112
Figura A.57 Img_7 - Trimap automático . . . 113
Figura A.58 Img_7 - Composição obtida com o Bayesian Matting a partir do Trimap manual . . . 113
Figura A.59 Img_7 - Composição obtida com o Bayesian Matting a partir do Trimap automático . . . 114
Figura A.60 Img_7 - Composição obtida com o Robust Matting a partir do Trimap manual . . . 114
Figura A.61 Img_7 - Composição obtida com o Robust Matting a partir do Trimap automático . . . 115
Figura A.62 Img_7 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual . . . 115
Figura A.63 Img_7 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático . . . 116
Figura A.64 Img_8 - Imagem capturada pela câmera RGB . . . 116
Figura A.65 Img_8 - Trimap manual . . . 117
Figura A.66 Img_8 - Trimap automático . . . 117
Figura A.67 Img_8 - Composição obtida com o Bayesian Matting a partir do Trimap manual . . . 118
Figura A.68 Img_8 - Composição obtida com o Bayesian Matting a partir do Trimap automático . . . 118
Figura A.69 Img_8 - Composição obtida com o Robust Matting a partir do Trimap manual . . . 119

Figura A.70 Img_8 - Composição obtida com o Robust Matting a partir do Trimap automático . . . 119
Figura A.71 Img_8 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual . . . 120
Figura A.72 Img_8 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático . . . 120
Figura A.73 Img_9 - Imagem capturada pela câmera RGB . . . 121
Figura A.74 Img_9 - Trimap manual . . . 121
Figura A.75 Img_9 - Trimap automático . . . 122
Figura A.76 Img_9 - Composição obtida com o Bayesian Matting a partir do Trimap manual . . . 122
Figura A.77 Img_9 - Composição obtida com o Bayesian Matting a partir do Trimap automático . . . 123
Figura A.78 Img_9 - Composição obtida com o Robust Matting a partir do Trimap manual . . . 123
Figura A.79 Img_9 - Composição obtida com o Robust Matting a partir do Trimap automático . . . 124
Figura A.80 Img_9 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual . . . 124
Figura A.81 Img_9 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático . . . 125

Lista de Tabelas

7.1 Tempo de processamento dos métodos de *matting* digital em segundos, a partir do *Kinect*. 70

7.2 *Trimap* Automático / *Trimap* Manual (%) 70

7.3 *Trimap* Manual - *Trimap* Automático (em segundos). 70

7.4 Comparação de qualidade entre as imagens geradas pelos métodos de *matting* digital com *trimap* automático e *trimap* manual, a partir do *Kinect*. 71

7.5 Tempo de processamento dos métodos de *matting* digital em segundos, a partir do *Chromakey*. 76

7.6 *Trimap* Automático / *Trimap* Manual (%) 77

7.7 *Trimap* Manual - *Trimap* Automático (em segundos). 77

7.8 Comparação de qualidade entre as imagens geradas pelos métodos de *matting* digital com *trimap* automático e *trimap* manual. 78

Lista de Abreviaturas

2D Duas dimensões
3D Três dimensões
CMOS *Complementary metal-oxide-semiconductor*
FPS *Frames per second*
fullHD *Full High Definition*
Gb *Giga Byte*
HD *High Definition*
IR Infravermelho
RGB *Red, Green, Blue*
RGBA *Red, Green, Blue, Alpha*
SDK *Software Development Kit*
SSIM *Strutural Similarity*
TOF *Time Of Flight*
UIQI *Universal Image Quality Index*
VGA *Video Graphics Array*

CAPÍTULO 1

Introdução

1.1 Contextualização

Muita das imagens exibidas no vídeo, em transmissões de televisão ou em cenas de filmes, são geradas pela combinação de duas camadas: uma contendo os objetos de primeiro plano (*foreground*) e outra contendo o plano de fundo (*background*). Para a realização desta composição, há mais de duas décadas, vem se estudando técnicas de *Matting* Digital ((CHUANG et al., 2001),(WANG; COHEN, 2007b)).

A primeira etapa do *matting* digital consiste, geralmente, em segmentar e extrair os objetos em primeiro plano do plano de fundo, e para tal deve-se levar em consideração o nível de transparência dos objetos. O elemento canal alfa (α) foi desenvolvido para preservar o nível de transparência dos objetos a serem extraídos. Em termos matemáticos, o canal alfa possibilita controlar a interpolação linear entre as cores das duas camadas (primeiro plano e o plano de fundo) para cada *pixel*. O canal alfa contém valores que variam em uma escala de tons de cinza. Aos *pixels* com alta transparência, os quais não devem ser exibidos, deve ser atribuído o valor alfa igual a zero, em contrapartida o valor máximo deve ser atribuído aos *pixels* opacos, os quais devem ser preservados. Os *pixels* com algum nível de transparência devem conter valores intermediários de alfa (PORTER; DUFF, 1984).

Tanto a indústria da televisão quanto na indústria do cinema dispõem de recursos para utilizar uma camada de primeiro plano isolada, criando a necessidade de retirá-la de seu contexto original. Esta separação, conhecida como extração de chave, tem como objetivo uma futura combinação com camadas de planos de fundo originadas de diversas fontes, que podem ser filmagens realizadas em outros locais ou imagens produzidas em computadores.

A técnica do *chromakey*, há algum tempo vem sendo muito utilizada no cinema e na

televisão, inclusive com transmissões ao vivo. Seu princípio baseia-se no reconhecimento de partes da imagem, em que se encontra uma determinada cor, denominada cor chave. Uma máscara é então gerada para selecionar quais *pixels* devem permanecer e quais devem ser substituídos por *pixels* de outra imagem, utilizada como plano de fundo (BERGH; LALIOTI, 1999).

O grau de complexidade do problema é melhor representado quando, por exemplo, além da transparência dos objetos a serem extraídos, estes possuírem uma superfície irregular contendo pêlos, cabelos, plumas entre outros, ou ainda, existirem sombras e a reconstrução tridimensional dessa sombra no novo plano de fundo.

Conceitualmente, a operação de *chromakey* é simples, mas apenas a comparação dos *pixels* da imagem de entrada com a cor chave, selecionada automaticamente ou por meio do operador, pode conduzir a composições grosseiras e pouco naturais.

Abordagens muito simplificadas produzem uma série de imperfeições e, por esse motivo, os sistemas modernos que utilizam o *chromakey* aplicam algoritmos avançados para preservar detalhes do objeto em primeiro plano.

Alguns problemas em potencial são comuns a técnica, caso nenhum controle adicional seja implementado no sistema. Podem-se observar imperfeições nas bordas ao redor dos objetos extraídos do primeiro plano. Sombras sobre o plano de fundo podem ser reconhecidas como elementos em primeiro plano, e a câmera de vídeo deve permanecer imóvel durante o processo para manter a integridade entre as camadas de primeiro e segundo plano (SMITH; BLINN, 1996).

Diversos algoritmos encontrados na literatura fazem uso de um recurso conhecido como *trimap*, para a obtenção de mapas de transparência de maior qualidade. Normalmente o *trimap* é realizado com marcações feitas pelo usuário, informando o que é primeiro plano, plano de fundo, e a área indeterminada (fronteira entre o primeiro plano e plano de fundo). O *trimap* reduz a dimensão do espaço de solução do problema *matting*, o que leva os algoritmos de *matting* a gerar os resultados desejados pelo usuário. (WANG; COHEN, 2007b).

A partir de dispositivos de luz estruturada, os quais possibilitam a criação do mapa da profundidade da imagem de entrada, é possível utilizar o mapa de profundidade como uma informação adicional no processo do *matting* digital contribuindo para a composição da imagem final. O mapa de profundidade é uma representação em tons de cinza da imagem de entrada, onde os *pixels* mais claros estão mais próximos da câmera e os *pixels* mais escuros estão mais distantes da câmera. Um desses dispositivos de luz estruturada que possibilita o mapa de profundidade é o *Kinect* desenvolvido pela Microsoft para o videogame Xbox 360.

1.2 Objetivos do Trabalho

Objetivo geral

- Desenvolver e testar métodos para a geração automática de *trimaps*, que possam ser utilizados como parte do processo de *Matting* Digital, aplicáveis em imagens e vídeos. Para o desenvolvimento destes métodos assume-se a existência de informação extra sobre a imagem, tais como, mapa de profundidade e/ou fundo de cor arbitrária.

Objetivo específicos

- Levantar as principais técnicas de *matting* digital, em especial aquelas que se baseiam em informações extras;
- Estudar as funcionalidades do equipamento de luz estruturada conhecido como *Kinect* da Microsoft, pertinentes ao desenvolvimento deste projeto;
- Estruturar e implementar métodos para geração automática de *trimaps* para imagens;
- Verificar a potencialidade da aplicação dos métodos para vídeo; e
- Comparar as implementações dos *trimaps* gerados automaticamente com aqueles gerados manualmente.

1.3 Organização do Trabalho

Os capítulos deste trabalho foram organizados da seguinte forma:

- O capítulo 2 apresenta os principais conceitos utilizados para o desenvolvimento das técnicas de *matting* digital abordadas neste trabalho.
- O capítulo 3 apresenta os principais conceitos sobre *Trimap*.
- O capítulo 4 descreve os componentes do dispositivo de luz estruturada, o *Kinect*.
- O capítulo 5 apresenta os métodos desenvolvidos para a automatização de *trimap* para imagens e vídeo.
- O capítulo 6 descreve as implementações dos métodos desenvolvidos.
- O capítulo 7 descreve os experimentos, os resultados obtidos durante a realização do projeto, bem como a análise dos mesmos.
- O capítulo 8 apresenta as conclusões do projeto, e elenca algumas sugestões para trabalhos futuros.

CAPÍTULO 2

Matting Digital

Este capítulo apresenta definições e conceitos de *matting* digital juntamente com as algumas técnicas existentes, as quais podem, ou não envolver a utilização de informações extras e *trimaps*.

2.1 Considerações Iniciais

A demanda por maior realismo na geração de efeitos especiais em imagens e vídeo a fim de atender aplicações da indústria cinematográfica, televisão, dentre outras, requer o desenvolvimento de técnicas para a composição de imagens com alta qualidade. Este capítulo apresenta conceitos e terminologias inerentes ao *matting* digital, além de um breve histórico da evolução das técnicas de composição de imagens.

2.2 Definições e Conceitos Iniciais

A necessidade da indústria cinematográfica em combinar, de forma convincente, imagens filmadas em locais e momentos diferentes em uma única faixa de filme, impulsionou o desenvolvimento da técnica de composição de imagens.

O *matting* e a composição de imagens teve sua origem nas indústrias fotográficas e cinematográficas. Um dos primeiros exemplos de composição de imagens foi a impressão massiva, "*The Two Ways of Life*", criada pelo fotógrafo Oscar Rejlander em 1857 por meio da combinação seletiva de 32 negativos diferentes (BRINKMANN, 1999).

A técnica da "revelação dupla" foi uma das primeiras soluções para a composição de imagens, onde partes do filme eram preservadas no momento da captura. Assim com

um novo processo de filmagem, as áreas protegidas na primeira captura recebem novos elementos (RICKITT, 2000).

Essas combinações de filmes foram produzidas por meio de impressoras ópticas, especialmente desenvolvidas para esta finalidade. Nascia então a arte da composição óptica. No processo de composição óptica são necessários três faixas de filme para gerar a composição (BIEDNY; MOODY, 1998):

- Faixa do primeiro plano: o filme com os elementos em primeiro plano que serão combinados com um novo plano de fundo.
- *Matte*: o filme monocromático o qual determina a opacidade da faixa do primeiro plano. A região ao redor dos objetos a serem extraídos é preta para bloquear o plano de fundo indesejável da faixa do primeiro plano. E a região interna dos objetos em primeiro plano é transparente.
- Faixa do plano de fundo: o filme com novo plano de fundo a ser combinado com os elementos da faixa do primeiro plano.

Durante o processo de composição, primeiramente o filme do plano de fundo é exposto através da negativa do *matte* (*male matte*) em uma nova faixa de filme, a fim de manter uma cópia do plano de fundo com uma área preservada. Esta área tem a mesma forma e tamanho do elemento de primeiro plano. A nova faixa de filme é, então, rebobinada e reexposta à faixa de primeiro plano juntamente com o *matte* (*female matte*) (RICKITT, 2000).

Assim a nova exposição ocorre somente onde o *matte* é transparente gerando um novo filme com a composição final. Com as impressoras ópticas, a composição se tornou uma tarefa simples. No entanto, a dificuldade em gerar *mattes* apropriados impulsionou o desenvolvimento de diversas técnicas (CHUANG, 2004).

A Figura 2.1 ilustra o processo de composição óptica.

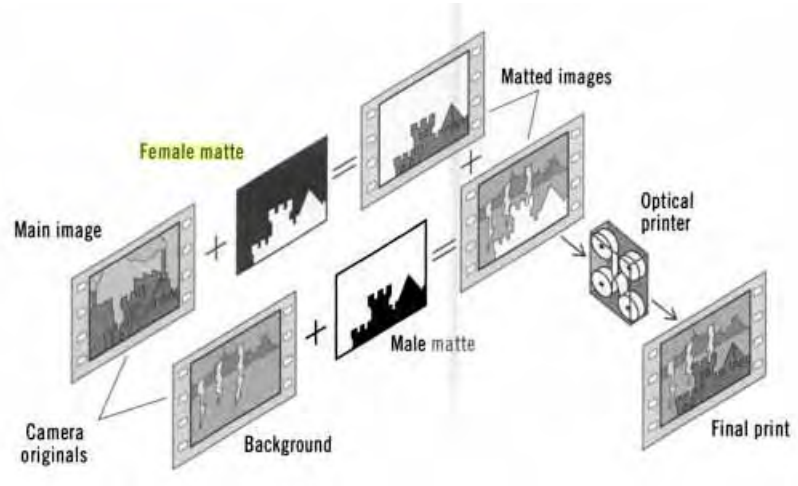


Figura 2.1: *Composição óptica. Fonte: (MONACO, 2000)*

Após esse breve histórico da composição de imagens ópticas é possível explicar o termo *matting*, sendo este oriundo da palavra *matte*.

O *matting* digital é o processo de composição de imagens digitais, podendo ser divido em duas fases. A primeira fase é a extração dos objetos em primeiro plano, e a segunda é a combinação desses objetos em primeiro plano com um novo plano de fundo (WANG; COHEN, 2007a).

O *matting* digital pode parecer um problema simples, porém a extração dos objetos em primeiro plano é complexa quando estes objetos possuem uma superfície irregular, tais como cabelo e pêlos, ou serem extremamente texturizados, ou ainda possuírem alguma transparência.

Com o intuito de facilitar a representação da transparência dos objetos a serem extraídos, foi elaborado o elemento denominado canal alfa (α) (PORTER; DUFF, 1984). Esse elemento é utilizado para representar níveis de transparência dos *pixels* da imagem em primeiro plano, possibilitando a extração dos elementos de interesse e ainda a preservação da transparência dos objetos. Com um enfoque mais matemático, o canal alfa possibilita controlar a interpolação linear entre as cores das duas camadas (primeiro plano e o fundo) para cada *pixel*.

O canal alfa possui valores que variam em uma escala de tons de cinza (PORTER; DUFF, 1984). Aos *pixels* com alta transparência, os quais não devem ser exibidos, deve ser atribuído o valor alfa igual a zero, em contrapartida o valor máximo deve ser atribuído aos *pixels* opacos, os quais devem ser preservados. Os *pixels* com algum nível de transparência devem conter valores intermediários de alfa (PORTER; DUFF, 1984).

A utilização do canal alfa proporciona a criação de imagens digitais com alto grau de complexidade e, por esse motivo, a maioria dos métodos de composição é baseada na "Equação da Composição" (Eq. 2-1):

$$C = \alpha F + (1 - \alpha)B \tag{2-1}$$

Onde:

C é a cor resultante representada na composição;
F é a cor do elemento em primeiro plano (*foreground*);
B é a cor do novo plano de fundo (*background*); e
α é a componente de transparência do *pixel*.

Com uma imagem de entrada, para cada *pixel* são conhecidos os vetores tridimensionais de cor *Red, Green, Blue* (RGB), porém são desconhecidos os valores das variáveis α, B e F. A maioria dos métodos de composição de imagens depende de conhecimentos prévios para restringir o problema, possibilitando uma boa estimativa dessas variáveis. Uma vez estimadas corretamente essas variáveis, o primeiro plano pode ser suavemente integrado ao um novo plano de fundo, através da substituição do fundo original B por B' da nova imagem de plano de fundo na equação (Eq. 2-1), como mostra a Figura 2.2.

(a) (b) (c)

Figura 2.2: *Exemplo de matting com o uso da tela azul de fundo, onde (a) é o objeto à ser extraído, (b) objeto extraído e (c) composição final. Fonte: (SMITH; BLINN, 1996)*

Analisando o *matting* digital com enfoque mais voltado para o processamento de imagens, pode-se restringir os valores de alfa para zero (0) e um (1), e o problema do *matting* torna-se o problema clássico da segmentação binária da imagem ou vídeo.

Assim cada *pixel* pertence plenamente ao primeiro plano ou ao plano de fundo. Porém a dificuldade está em reconhecer os objetos em primeiro plano por meio de uma imagem de duas dimensões. A fim de facilitar essa segmentação, algumas soluções encontradas na literatura utilizam uma imagem adicional de entrada com marcações manuais e diferentes para os objetos em primeiro plano e o plano de fundo. Essas marcações têm por finalidade gerar um mapa triplo (*trimap*). Com a informação adicional do *trimap*, o problema para identificar se um *pixel* pertence ou não ao primeiro plano é reduzido, podendo auxiliar os algoritmos de *matting* que se utilizem desta técnica.

2.3 Os problemas do *matting* digital

Como mencionado anteriormente, é fácil perceber a dificuldade da extração de objetos em primeiro plano de uma imagem de duas dimensões, quando esses objetos têm alguma transparência e ainda refletem outros objetos em sua superfície. A Figura 2.3 ilustra a complexidade das composições deste tipo.

Figura 2.3: *Composição de nove objetos. Fonte: (SMITH; BLINN, 1996)*

Além da transparência, outros problemas dificultam a extração dos objetos em primeiro plano, a superfície irregular do objeto contendo pêlos, cabelos, plumas entre outros, como mostra a Figura 2.4.

Figura 2.4: *Exemplo de objeto com superfície irregular e com sombra. Fonte: (MONTABONE, 2010)*

Outro problema, com relação ao *matting*, ocorre quando há necessidade em preservar a sombra do objeto em primeiro plano e fazer a composição dessa sombra com o novo plano de fundo. A dificuldade na composição da sombra com o novo plano de fundo, está no posicionamento da sombra sendo fiel ao relevo do novo plano de fundo. Ou seja, o novo plano de fundo é uma imagem 2D, porém deve-se considerar sua projeção 3D. Este exemplo é ilustrado na Figura (2.5).

(a) (b) (c)

Figura 2.5: *Exemplo de matting considerando a sombra do objeto em primeiro plano, onde (a) imagem entrada, (b) plano de fundo, (c) composição final preservando a sombra do objeto. Fonte: (CHUANG, 2004)*

2.4 *Matting* com informação extra

Informações adicionais que facilitem a obtenção do *matte* podem auxiliar a solução final. A informação extra mais clássica e frequentemente utilizada na indústria cinemato-

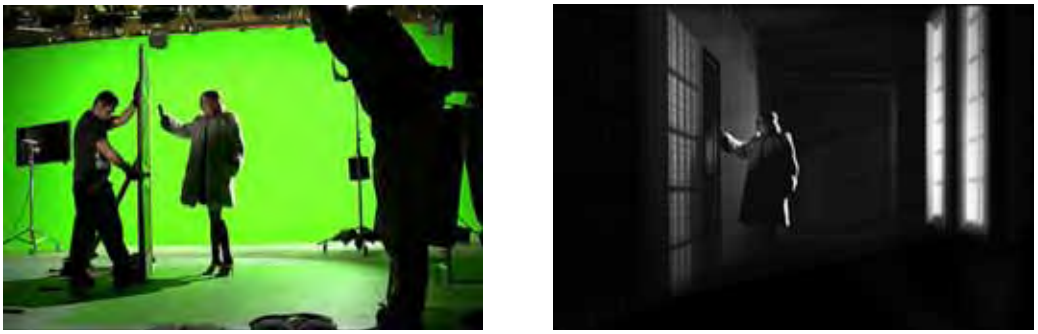
gráfica e da televisão, é o *matting* de tela azul (*Blue Screen Matting*) (SMITH; BLINN, 1996).

Neste caso em especial, o objeto em primeiro plano é filmado a frente de um plano de fundo com cor uniforme ou quase uniforme. Essa coloração de fundo normalmente é azul, no entanto outras cores também foram utilizadas como verde e amarelo.

Serão apresentadas, nas subseções a seguir, técnicas de *matting* digital que utilizam informação extra.

2.4.1 *Chromakey*

A técnica do *chromakey*, há algum tempo vem sendo muito utilizada no cinema e na televisão, inclusive com transmissões ao vivo. Seu princípio baseia-se no reconhecimento de partes da imagem, em que se encontra uma determinada cor, denominada cor chave. Uma máscara é então gerada para selecionar quais *pixels* devem permanecer e quais devem ser substituídos por *pixels* de outra imagem, utilizada como plano de fundo (BERGH; LALIOTI, 1999). A Figura 2.6 ilustra a técnica mencionada.



(a) (b)

Figura 2.6: *Exemplo de chromakey, onde (a) imagem original, (b) composição final. Fonte: (FOSTER, 2010)*

Este método faz a análise da imagem a partir das cores dos *pixels*, onde cada *pixel* é analisado separadamente. Para diminuir a quantidade de processamentos do algoritmo uma simplificação do cálculo de distância euclidiana (Eq.2-2), onde as componentes das cores são representadas pelos eixos no sistema cartesiano, como mostrado na Figura 2.7.

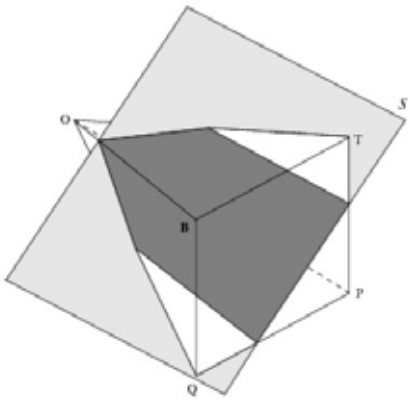


Figura 2.7: *Cubo RGB com a intersecção do plano S com a pirâmide de inclinação OBTPQ. Fonte: (BERGH; LALIOTI, 1999).*

$$D = B - R - 2G \tag{2-2}$$

Onde:

D caracteriza a proximidade entre a cor do *pixel* com o verde;
B componente azul do *pixel* no formato RGB;
R componente vermelho do *pixel* no formato RGB; e
G componente verde do *pixel* no formato RGB (cor do *background*).

Para este método o plano de fundo deve ser controlado e possuir uma cor chave uniforme (que deve ser verde ou azul). Com isso, para todos os *pixels* da imagem é calculada a equação (Eq. 2-2), e o resultado é comparado com um valor de distância pré-definido. Caso o valor do resultado da equação (Eq. 2-2) para o *pixel* em questão for superior ao valor pré-definido, o *pixel* é trocado pelo *pixel* correspondente do novo plano de fundo, uma vez que a distância é maior. Caso contrário o *pixel* não é trocado.

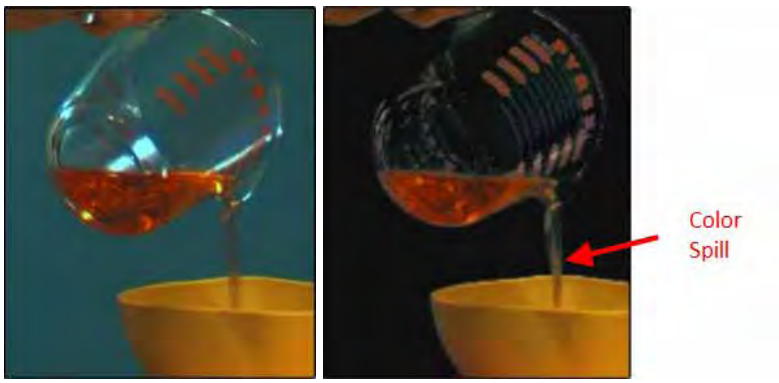
Cabe destacar que a precisão do método está diretamente relacionada com a uniformidade da cor do plano de fundo e sua iluminação, assim quanto menor for a variação RGB dos *pixels* do plano de fundo, mais preciso e eficiente será determinar o valor da distância. Outro fator relevante é o objeto em primeiro plano a ser recortado, não poder possuir tons da cor do fundo arbitrário.

Em outros termos, a operação de *chromakey* é simples, mas apenas a comparação dos *pixels* da imagem de entrada com a cor arbitrária, selecionada automaticamente ou por meio do operador, pode conduzir a composições grosseiras e pouco naturais.

Aplicações muito simplificadas produzem uma série de imperfeições e, por esse motivo, os sistemas modernos que utilizam o *chromakey* aplicam algoritmos avançados para preservar detalhes do objeto em primeiro plano.

Alguns problemas em potencial são comuns a técnica, caso nenhum controle adicional seja implementado no sistema. Podem-se observar imperfeições nas bordas ao redor dos objetos extraídos do primeiro plano. Sombras sobre o plano de fundo podem ser confundidas como elementos em primeiro plano, e a câmera de vídeo deve permanecer imóvel durante o processo para manter a integridade entre as camadas de primeiro e segundo plano (SMITH; BLINN, 1996).

Uma outra dificuldade a ser contornada, e que não é restrita aos planos de fundo de coloração constante, é o *color spill*. Esse é o problema da contaminação do objeto em primeiro plano com as cores do plano de fundo conforme ilustrado na Figura 2.8(b).



(a) *Imagem de entrada* (b) *Composição final*

Figura 2.8: *Exemplo do problema do color spill, onde (a) imagem original, (b) composição final. (MCGUIRE; MATUSIK; YERAZUNIS, 2006)*

O *color spill* ocorre nas bordas dos objetos em primeiro plano, onde os *pixels* acumulam a iluminação de ambas as camadas. Ou em casos de objetos transparentes em primeiro plano em que luzes refletidas os atravessam e retornam à câmera. São mais visíveis em fundos com cor verde, uma vez que estes são mais reflexivos.

Esta contaminação provoca a necessidade da remoção dessas impurezas, antes da composição do objeto em primeiro plano com o novo plano de fundo (GVILI et al., 2003).

2.4.2 *Flash matting*

O *flash matting* é outra técnica de *matting* que utiliza informação extra. Ao invés de ter apenas uma imagem de entrada, essa abordagem conta com a vantagem de um par de imagens de entrada, sendo uma com *flash* e a outra sem (SUN et al., 2006). A Figura 2.9 ilustra a técnica mencionada.

(a) *Imagem com flash* (b) *Imagem sem flash)*

(c) *Matte* (d) *Composição final)*

Figura 2.9: *Exemplo do Flash Matting. (SUN et al., 2006)*

A técnica foi desenvolvida inspirada na simples observação da diferença entre imagens com e sem *flash*, uma vez que essa diferença é bem visível nos objetos em primeiro plano, quando o plano de fundo está suficientemente distante.

Para restringir a dimensão espacial do problema do *flashing matting*, um *trimap* é utilizado como outra informação adicional. Este *trimap* é criado de forma semelhante ao método dos dois passos de Canny (CANNY, 1986). Este método é um processo de detecção de bordas que baseia-se em dois critérios, os de detecção e localização.

A Figura 2.10 mostra o resultado do *flashing matting* gerado a partir de um *trimap* automático, onde o objeto em primeiro plano é uma planta com alto grau de complexidade de extração.

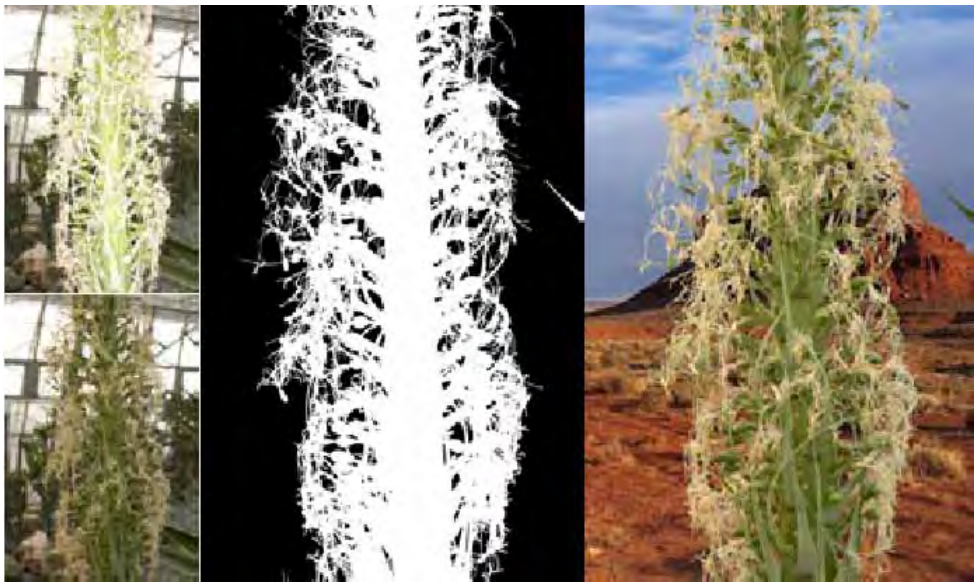


Figura 2.10: *Exemplo do flash matting para um objeto complexo.*
Fonte: (SUN et al., 2006)

Os autores informam que na maioria dos experimentos, o usuário precisa ajustar de forma interativa um limiar T (para detecção do primeiro plano) e a dilatação da região inderteminada (fronteira entre o primeiro plano e o plano de fundo). E para os casos extremamente complexos, há a necessidade das marcações diferenciadas para o primeiro plano e o plano de fundo.

Esta abordagem mostrou ser eficaz mesmo com as cenas em que o primeiro plano e o plano de fundo são similares, ou ainda quando o plano de fundo altamente texturizado. Os resultados, desenvolvidos pelos autores, envolvendo uma variedade de cenas complexas em ambientes fechados e abertos obteve *mattes* de alta qualidade.

2.4.3 *Matting* com Vetor de Câmeras

Utilizando um vetor de câmeras para a extração com alta qualidade dos objetos em primeiro plano (*mattes*) de uma sequencia de vídeo foi proposta por (JOSHI; MATUSIK; AVIDAN, 2006). Esse sistema de câmeras em paralelo é exibido na Figura 2.11.

Figura 2.11: *Vetor com 8 camêras e sistema em tempo real. Matte gerado com um quarto da resolução VGA. Fonte: (JOSHI; MATUSIK; AVIDAN, 2006)*

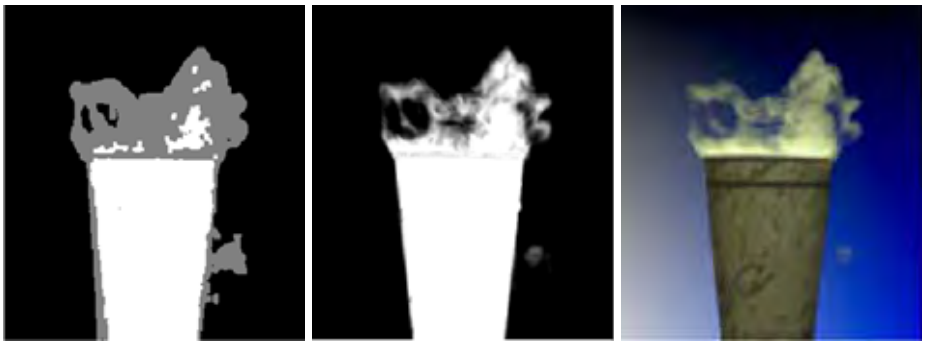
O sistema utiliza cenas naturais para calcular o *matte*, simulando o diafragma de uma câmera fotográfica usual. O diafragma tem um movimento de abertura e fechamento circular, e quanto maior for essa abertura maior será a quantidade de luz a atravessar o diafragma e menor será o campo de profundidade, e quanto menor for a abertura menor será a quantidade de luz a atravessar o diafragma e maior será o campo de profundidade. Esse diafragma artificial é na verdade o vetor de câmeras, assim a abertura determina a quantidade de câmeras que serão utilizadas.

Com base na técnica da parametrização dinâmica, que permite ao usuário modificar dinamicamente o foco e o campo de profundidade (ISAKSEN; MCMILLAN; GORTLER, 2000).

Os autores afirmam que o sistema é bem eficiente para a extração de objetos complexos como mostra a Figura 2.12, como objetos transparentes, fluidos e fumaça, e o tempo de processamento por *pixel* é linear em relação ao número de câmeras. Ainda permite a extração do *matte* quase em tempo real (JOSHI; MATUSIK; AVIDAN, 2006).

(a) *Imagem de entrada*

(b) *imagem com a variação das refocadas sintéticas da imagem (região mais escura onde houve menor variação)*



(c) *Trimap gerado a partir da variação da imagem*

(d) *Matte*

(e) *Composição final*

Figura 2.12: *Vídeo matting natural. (JOSHI; MATUSIK; AVIDAN, 2006)*

2.4.4 Câmeras *time of flight*

Recentemente, as câmeras TOF (*time of flight*) começaram atrair a atenção de muitos pesquisadores. Essas câmeras são sensores ativos que determinam, além da cor de cada *pixel*, também o valor da profundidade. Essa profundidade é calculada medindo o tempo em que a luz infravermelha demora para viajar da câmera até o objeto e retornar a câmera, daí vem o nome tempo de voo (*time of flight*).

Com a estimativa independente da profundidade por *pixel*, o mapa de profundidade pode possuir ruídos espacial e temporal. Esses ruídos são dependentes do conteúdo e por isso é difícil sua remoção através de filtros clássicos.

Mas o problema mais crítico ocorre quando existem objetos no plano de fundo que estão próximos ao primeiro plano, podendo gerar uma falsa segmentação do primeiro plano.

A abordagem *TofCut* (WANG et al., 2010) baseia-se no fato das câmeras TOF ainda não terem sido amplamente utilizadas nas aplicações de composição de imagens. Desenvolveram um algoritmo de segmentação em tempo real com a utilização da câmera TOF. Para os testes deste algoritmo, os autores utilizaram uma câmera ZCam da empresa 3DV Systems, com resolução de 320x240 tanto para o RGB quanto para a profundidade e uma taxa de 30 quadros por segundo. O algoritmo desenvolvido atingiu uma taxa de saída de 15 quadros por segundo. A Figura 2.13 exibe o resultado desta abordagem.

(a) *Câmera ZCam* (b) *Imagem de profundidade*

(c) *Imagem colorida* (d) *Resultado da segmentação*

Figura 2.13: *Uso da câmera TOF para a extrair o primeiro plano do plano de fundo. (WANG et al., 2010)*

2.5 Considerações Finais

A partir do estudo mostrado neste capítulo, verificou-se o processo do *matting* digital juntamente com as dificuldades inerentes ao processo.

A técnica de *Flash Matting* é interessante e eficaz no processo de *matting* digital para imagens estáticas, porém não pode ser aplicada em vídeo.

A utilização do vetor de câmeras é eficiente na extração dos objetos em primeiro plano, porém existe a necessidade da utilização do *hardware* específico.

A câmeras *time of flight* possibilitam a criação do mapa de profundidade da imagem, e podem vir a contribuir no processo de *matting* digital, porém atualmente ainda possuem uma baixa resolução e um alto custo.

A importância da informação extra no processo de *matting* digital, tendo em vista a gama de problemas inerentes ao processo de *matting* digital . Assim o próximo capítulo será demonstrado o *trimap*, bem como o métodos de *matting* que o utilizam.

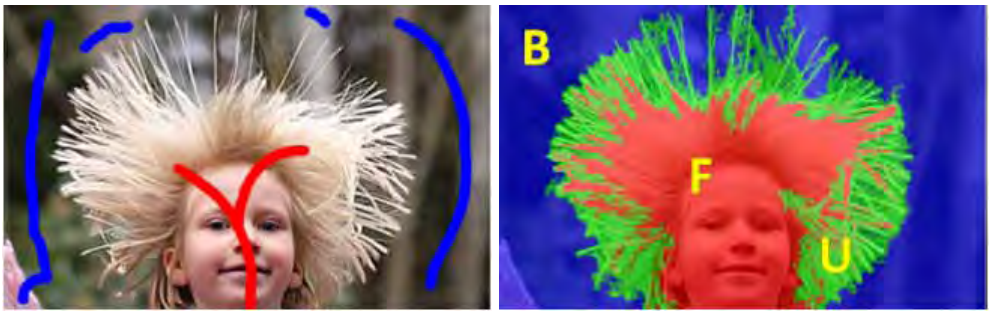
CAPÍTULO 3

Trimap

Neste capítulo será apresentada a informação extra para o processo de *matting* digital, o *Trimap*.

3.1 Considerações Iniciais

Com intuito de aprimorar a extração dos objetos em primeiro plano, algumas abordagens de *matting* digital utilizam como informação extra a imagem de entrada segmentada em três regiões: definindo o primeiro plano, o plano de fundo, e a região desconhecida (fronteira entre o primeiro plano e plano de fundo). O conjunto dessas três regiões é chamado de *trimap* conforme ilustra a Figura 3.1(b) (WANG; COHEN, 2007a).



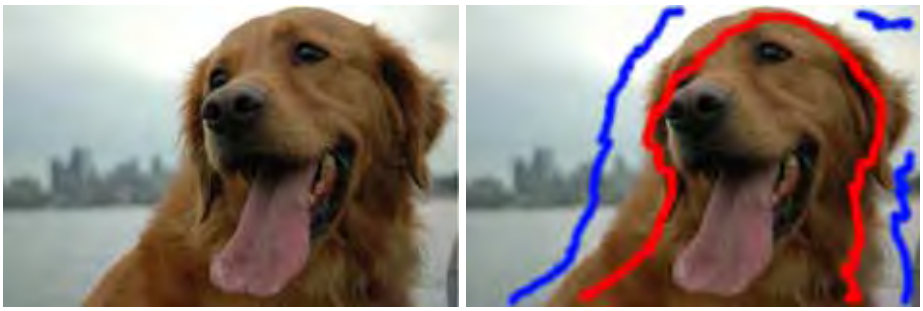
(a) *Imagem de entrada com as marcações realizadas pelo usuário*

(b) *Trimap, sendo a região B o plano de fundo, região F é o primeiro plano e a região U é a região desconhecida (região indeterminada)*

Figura 3.1: *Exemplo do trimap. Fonte: (RHEMANN et al., 2008)*

Normalmente o *trimap* é gerado de forma manual, a partir de marcações realizadas pelo usuário, feitas para diferenciar as regiões do plano de fundo e o primeiro plano. A

Figura 3.2(b) exibe tais marcações.



(a) *Imagem de entrada* (b) *Marcações do usuário)*

(c) *Trimap*

Figura 3.2: *Exemplo do trimap. Fonte: (HU, 2010)*

A partir dessas marcações realizadas pelo usuário , pode-se classificar o *trimap* em dois tipos principais, o *stroke trimap* utilizado para diminuir a intervenção do usuário no momento de especificar o *trimap*, e o *filled trimap* executado para gerar resultados mais precisos para os algoritmos de *matting* digital (LIMBERGER, 2011).

3.1.1 *Stroke Trimap*

O modelo *stroke trimap* consiste em desenhar traços com a finalidade de limitar o plano de fundo e o objeto em primeiro plano da uma imagem de maneira simples, mas com pouca precisão.

Com essa metodologia os *pixels* que não foram marcados como primeiro plano ou plano de fundo são tratados como área indeterminada, por isso, os algoritmos que possam vir a utilizar este tipo de *trimap* necessitem de muitos traços adicionais para gerar resultados satisfatórios, além de utilizarem maior tempo de processamento tendo

em vista a grande área indeterminada do *trimap*. Um exemplo de *stroke trimap* pode ser representado pela Figura 3.2(b).

3.1.2 *Filled Trimap*

O tipo *filled trimap* é utilizado para marcar totalmente o plano de fundo e o primeiro plano da imagem, deixando uma pequena quantidade de *pixels* para a região indeterminada. Deste modo, toda a imagem é pintada com uma das três cores: preta (região do plano de fundo), cinza (região desconhecida) ou branca (região de primeiro plano).

É claro perceber que o *filled trimaps* geram um resultado mais preciso que o *stroke trimap* e um menor tempo de processamento para os algoritmos de *matting* digital. Porém o tempo de interação do usuário é maior neste caso. Um exemplo de *filled trimap* pode ser visualizado na Figura 3.2(c).

3.2 *Trimap* automático

Uma aproximação automática do *trimap* sem marcação pelo usuário na imagem de entrada, foi proposta no algoritmo *Spectral matting* (LEVIN; RAV-ACHA; LISCHINSKI, 2008). No entanto os autores concordam que a aproximação automática tem suas limitações, incluindo resultados errôneos para as imagens com um fundo altamente texturizados. Com isso, na prática, um *matting* de alta qualidade que faz uso do *trimap*, ainda necessita da especificação do primeiro plano e o plano fundo por meio da interação do usuário.

O presente trabalho não utiliza esta técnica, uma vez que utiliza como informação extra o mapa de profundidade gerado a partir de dispositivos de luz estruturada.

3.3 Métodos de *Matting* digital

Nas próximas subseções serão apresentados os métodos de *matting* digital, os quais podem utilizar o *trimap* como informação extra.

3.3.1 *Bayesian Matting*

Uma abordagem que utiliza o *trimap* como informação extra é o *Baeysian matting* desenvolvido por (CHUANG et al., 2001), semelhante ao algoritmo Ruzon e Tomasi (RU-

ZON; TOMASI, 2000), que faz uso de amostras de cor, gerando distribuições gaussianas, entretanto utiliza algumas modificações. A primeira delas é o uso de uma janela, esta que percorre a imagem de forma contínua, para definir as vizinhanças sobre as imagens de fundo e frente, gerando o *alphamap* .

O método utiliza novos conhecimentos e amostras do *background* para construir distribuições de cores, também são utilizadas as proximidades do *foreground*, do *background*, de modo que todos os *pixels* na vizinhança contribuam para formar as distribuições gaussianas do primeiro plano e plano de fundo, ver Figura 3.3

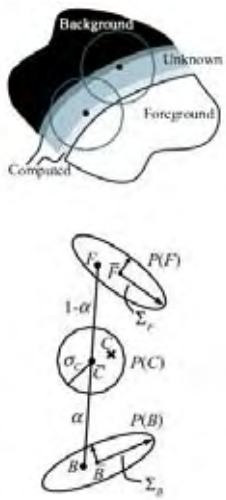


Figura 3.3: *Análise local do pixel desconhecido. Fonte: (CHUANG et al., 2001)*

As probabilidades são modeladas pela derivada da distribuição Gaussiana. Quando o *pixel* é atribuído para ambos os planos, é realizado um procedimento de otimização para o par de pontos plano de fundo e primeiro plano da imagem. O melhor resultado é escolhido pela máxima semelhança. O resultado pode ser mais preciso com a utilização de um *trimap* bem definido, caso contrário contrária os resultados podem ser grosseiros, como exibidos na Figura 3.4.

Figura 3.4: *Exemplo dos resultados do Bayesian matting. Fonte: (CHUANG et al., 2001)*

3.3.2 *Robust Matting*

Outro método que pode-se utilizar do *trimap* como informação extra é o *Robust matting*, que combina a amostragem otimizada em cores com o *matting affinity*, resultando em um sistema bem equilibrado, que é capaz de gerar resultados de alta qualidade, mantendo um razoável grau de robustez com diferentes entradas do usuário. O método utiliza uma função de energia, esta que é minimizada (WANG; COHEN, 2007b) . São criados nós virtuais, estes que representam o "fundo puro", ou a "frente pura", conforme é exibido na Figura 3.5, em que ΩF e ΩB representam os nós virtuais de *foreground* (F) e o *background* (B), respectivamente e a cor branca representa os pontos indeterminados. Estes nós ideais são utilizados para encontrar os valores de alfa. Essa metodologia é adotada, pois nem sempre as amostras de cor são totalmente confiáveis, dessa forma os resultados são melhores (WANG; COHEN, 2007a). Cabe salientar que os *pixels* de certeza do primeiro plano ou plano de fundo são marcados pelo usuário, por meio do *trimap* manual.

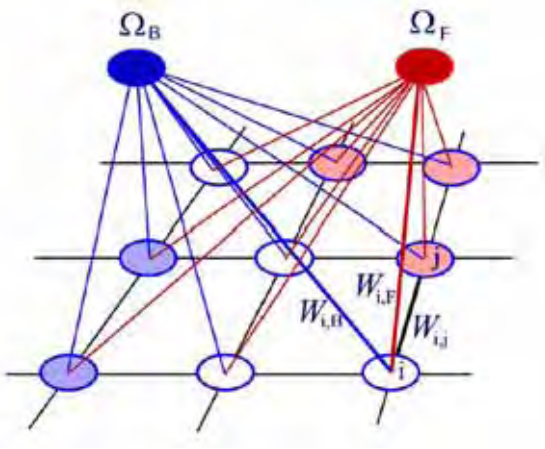


Figura 3.5: *Análise do pixel desconhecido. Fonte: (WANG; COHEN, 2007b)*

3.3.3 *Geodesic Matting*

O *Geodesic Matting* é outro método de *matting* digital que também utiliza o *trimap* como informação extra de entrada. Essa entrada é realizada por meio de marcações feitas pelo usuário, a partir das marcações são criados mapas tridimensionais, denominados mapas de cor, e estes são semelhantes ao mapa de um relevo. A partir deste mapa é calculada a distância geodésica dos *pixels* desconhecidos para os *pixels* conhecidos (pontos que foram marcados pelo usuário). Com essa distância é possível estimar um valor para alfa(α). A Figura 3.6 exibe resultados gerados a partir do *Geodesic Matting* (BAI; SAPIRO, 2009).

Figura 3.6: *Resultados obtidos com o Geodesic Matting. Fonte: (BAI; SAPIRO, 2009)*

3.3.4 *Geodesic* Modificado

O artigo (BAI; SAPIRO, 2009) sugere que o método seja executado duas vezes, com a finalidade de refinar o resultado. Na primeira vez que o programa é executado são utilizados rabiscos feitos manualmente e, a partir deles é gerado um *alphamap*. Com

ele são feitas marcações nos limites entre os planos (da frente e de fundo) e áreas de transparências, estes são considerados os rabiscos ideais, ou seja o mais próximo possível dos pontos indeterminados. Com estes novos rabiscos o método é executado novamente, para obter o resultado final (ARRUDA, 2012).

Esta abordagem utiliza o *trimap* como entrada, juntamente com a imagem a ser recortada, desta forma os rabiscos foram feitos a partir dos *pixels* que estão nos limites entre os planos (da frente e de fundo) e as áreas indeterminadas, gerando também amostras de cor nas áreas próximas as transparências. Além disso os valores de alpha são calculados apenas nas áreas inderteminadas do *trimap*, para diminuir o tempo de processamento (ARRUDA, 2012).

3.4 Considerações Finais

A precisão do *trimap* influencia diretamente o resultado do algoritmo, e em condições ideais, a região desconhecida no *trimap* deve conter apenas os *pixels* não identificados como pertencentes ao primeiro plano. Em outras palavras, a fronteira entre o primeiro plano e a região desconhecida deve ser a mais estreita possível, possibilitando a redução do tempo gasto para calcular o *matte*. Assim, fica explícita a dependência da precisão em que se encontra o *trimap*, da quantidade de variáveis desconhecidas para estimar, e o grau de conhecimento do primeiro plano e o plano de fundo. Porém, especificar precisamente um *trimap* requer uma habilidade e prática do usuário, especialmente para objetos com muitas regiões semitransparentes ou vazados. Portanto há uma dificuldade da utilização do *trimap* em imagens em tempo real, tendo em vista a necessidade da participação interativa do usuário.

CAPÍTULO 4

Kinect

Este capítulo será apresentado do dispositivo de luz estruturada, o *Kinect*, desenvolvido pela Microsoft para o videogame Xbox 360, e que foi utilizado neste trabalho.

4.1 Considerações Iniciais

Kinect é um dispositivo inovador, desenvolvido pela Microsoft, que possui uma câmera que capta a posição em três dimensões do ambiente à sua frente, que foi desenvolvido para ser utilizado com o videogame Xbox 360. Este equipamento permite ao usuário interagir com o Xbox 360 sem a necessidade de tocar em um controle, através de uma interface natural que é o próprio corpo humano. Através de gestos e comandos por voz, o jogador pode navegar em interfaces do videogame e jogar jogos que tenham compatibilidade com o dispositivo.

4.2 Componentes internos do *Kinect*

O *Kinect* é constituído por três componentes ópticos: um projetor laser infravermelho (IR) de baixa intensidade, uma câmera de infravermelho e uma câmara colorida de alta resolução (HD). Uma característica específica do *Kinect* é a combinação de um projetor (IR) e uma câmera (IR), ao invés das habituais duas câmeras para visão estéreo. A principal vantagem desta combinação é a redução do volume de processamento (KUHN, 2011). Para o áudio do *Kinect* existe um vetor de microfones, com a composição de quatro sensores possibilitando separar o som vindo de diversas direções. Conta também com um motor na base para movimentação vertical, que foi desenvolvido para ser utilizado com o videogame Xbox 360. A Figura 4.1 ilustra a parte interna do *Kinect*.

Figura 4.1: *Imagem da parte interna do Kinect Fonte: (KUHN, 2011)*

4.3 Mapa de profundidade

A câmera de profundidade é composta por um projetor que emite radiações infravermelhas invisíveis ao olho humano e um sensor CMOS monocromático que recebe os dados das radiações refletidas no ambiente. Deste modo, o *Kinect* fornece para o usuário uma imagem composta de pontos, onde cada ponto representa a distância real do aparelho até o primeiro objeto no qual as radiações infravermelhas foram refletidas. Esta tecnologia possui algumas limitações, como por exemplo, um limite da distância mínima e máxima que o sensor consegue alcançar. A Microsoft recomenda uma distância mínima de um metro, e distância máxima de cinco metros para os sensores funcionarem normalmente. A partir destes limiares os dados se tornam nulos ou pouco precisos. Para captura de todo corpo humano (dos pés à cabeça) a Microsoft aconselha a distância mínima de 1,8 metros. A luz estruturada no *Kinect* é utilizada para calcular a profundidade. Para isso o projetor infravermelho IR emite uma nuvem de pontos padrão, esta nuvem contém 633x495 pontos. A partir da distorção dos pontos sobre a superfície é possível calcular a distância entre cada ponto e o projetor (KUHN, 2011). A Figura 4.2 mostra o mapa de profundidade capturado pela câmera IR (Figura 4.2(a)), com a identificação do ser humano e destacado na cor amarela, já a imagem (Figura 4.2(b)) foi capturada pela câmera RGB sem nenhum tratamento.

(a) *Mapa de profundidade* (b) *Imagem da câmera RGB*

Figura 4.2: *Exemplo das imagens capturadas pelo Kinect*

4.4 *Real-Time Human Pose Recognition*

Os desenvolvedores do sistema do *Kinect* analisaram uma grande quantidade de dados de capturas de movimentos, em diversos cenários da vida real, utilizando uma máquina de aprendizado. Assim foi possível mapear os dados para modelos que representavam pessoas em fisionomias diferentes. Com estes dados modelados, os pesquisadores conseguiram ensinar o sistema a detectar milhares de tipos de poses humanas, mapeando-as para um esqueleto tridimensional (SHOTTON et al., 2011). Este algoritmo gerado a partir do aprendizado do sistema foi chamado de *Real-Time Human Pose Recognition*. A partir do esqueleto gerado pelo *Kinect*, foi possível a detecção de gestos e movimentos, alavancando assim, a gama de possibilidades de programas e jogos que pudessem ser gerados a partir deste dispositivo. Cabe lembrar que os dados referentes ao reconhecimento de jogadores e de seus esqueletos são processados diretamente no *hardware* do *Kinect*, sem a necessidade de um *software* específico para trazer esta inovação ao computador ou ao Xbox.

4.5 Pacote de Desenvolvimento de Software para o *Kinect*

O *Kinect for Windows* SDK é uma ferramenta que permite ao desenvolvedor fácil acesso às capacidades oferecidas pelo dispositivo *Kinect*. Este pacote de desenvolvimento inclui drivers, acesso aos dados do sensor de profundidade e da câmera colorida RGB, funções para movimentação da inclinação do *Kinect* ao solo, funções para rastreamento de corpos humanos, documentação e exemplos de aplicações. O SDK não fornece nenhum

algoritmo de visão computacional ou processamento de imagens para serem aplicados nos dados adquiridos pelos sensores do *Kinect*. Caso a aplicação do usuário necessite utilizar filtros ou algoritmos de segmentação ou reconhecimento de elementos, o desenvolvedor deverá adicionar outras bibliotecas ao seu projeto para suas devidas funcionalidades.

4.6 Considerações Finais

O *Kinect* é um dispositivo de baixo custo capaz de gerar um mapa de profundidade em tempo real, cabe lembrar que esse mapa de profundidade é de baixa resolução.

O *Kinect* possui limitações em relação à distância do objeto ao dispositivo, qualquer objeto deve estar no mínimo a 80cm do equipamento, caso esse objeto seja um ser humano a distância recomendada pela Microsoft é 180cm.

CAPÍTULO 5

Métodos de Automatização de *Trimap* para imagens e vídeo

Este capítulo apresenta os métodos para automatização de *trimap* propostos neste trabalho.

5.1 Considerações Iniciais

No processo de *matting* digital, a imagem de entrada é a imagem que contém os objetos a serem extraídos. Com essa imagem de entrada podem ser utilizados métodos de amostragem de cor ou afinidades entre *pixels* e uma vizinhança pequena para estimar o valor de alfa (α) para cada *pixel*. A partir da estimativa dos valores de alfa é obtido o *alphamap*.

Com as informações do *alphamap*, da imagem do novo plano de fundo e da imagem de entrada, para cada *pixel* são substituídas as respectivas variáveis na equação da composição (Eq. 2-1) resultando a cor do *pixel* da imagem composta, ou seja, o *pixel* resultante da imagem final. Assim a imagem composta é a imagem resultante do processo de *matting* digital.

Os processos de *matting* digital, conforme descrito no capítulo 2, podem utilizar informações extras, como o *trimap*, cujo objetivo é minimizar o esforço computacional para o cálculo da estimativa dos valores de alfa.

Com a utilização do *trimap* como informação extra no processo de *matting* digital, uma nova etapa na estrutura do processo é adicionada, onde o usuário realiza (normalmente de forma manual), marcações na imagem de entrada cujo objetivo é determinar *pixels* que são realmente do primeiro plano e do plano de fundo. A partir das marcações definidas pelo usuário é gerado o *trimap*.

Após a geração do *trimap*, é utilizada a mesma estrutura para obter a composição final, porém cabe ressaltar que o processo para a estimativa dos valores de alfa não será realizada em todos os *pixels* da imagem, ou seja, apenas para os *pixels* da região indeterminada do *trimap* serão estimados os valores de alfa.

Serão descritos, nas subseções a seguir, os dois métodos automáticos de geração de *trimap* desenvolvidos. O primeiro baseia-se no uso de mapa de profundidade e o segundo, na técnica de *chromakey*.

5.2 Método automático de *trimap* baseado em mapa de profundidade

A partir do mapa de profundidade gerado pelo dispositivo de luz estruturada, o *Kinect*, foi desenvolvido o método automático na geração do *trimap*, sem a necessidade da interação do usuário para a realização de marcações.

A fim de representar a estrutura do processo de *matting* digital com o *trimap* gerado de forma automática, será detalhada a seguir a estrutura da geração automática do *trimap*.

Com a utilização do dispositivo de luz estruturada (*Kinect*), é obtido o mapa de profundidade da imagem de entrada. A partir do mapa de profundidade gerado juntamente com a imagem de entrada obtida em um ambiente com cor de fundo arbitrária, é aplicada a técnica de *chromakey* na imagem de entrada a fim de confrontar e validar o mapa de profundidade obtido pelo dispositivo.

Cabe ressaltar os problemas inerentes ao dispositivo, uma vez que o mapa de profundidade não tem a mesma resolução da imagem de entrada, além do dispositivo não informar com precisão no mapa de profundidade a distância dos *pixels* ao redor dos objetos em primeiro plano.

Após a geração do *trimap* de forma automática, são estimados os valores de alfa para os *pixels* da região indeterminada do *trimap* a partir dos algoritmos *Bayesian matting* (CHUANG et al., 2001), *Robust matting* (WANG; COHEN, 2007b) e *Geodesic* modificado (ARRUDA, 2012). A seguir, partindo-se dos valores de alfa, da imagem de entrada, e do novo plano de fundo são obtidos as cores dos *pixels* da composição final através da equação de composição (Eq. 2-1).

5.3 Método automático de *trimap* baseado em *chromakey*

O segundo método automático para geração de *trimap*, proposto neste trabalho, refere-se ao uso apenas da característica de cor, ou seja, utilizou-se a técnica de *chromakey* proposta por (BERGH; LALIOTI, 1999) e modificada por (ARRUDA, 2012). Nesta abordagem não é utilizado o mapa de profundidade, não havendo necessidade do uso de dispositivo de luz estruturada (*Kinect*), mas sim, de um ambiente com fundo de cor arbitrária (cor chave).

Com o intuito de exemplificar a estrutura do processo de *matting* digital com o *trimap* gerado de forma automática a partir da técnica de *chromakey*, assim a seguir será descrita a estrutura do método, sendo destacada a geração automática do *trimap*.

Como no método anterior, considera-se imagem de entranda aquela que contém os objetos a serem extraídos capturados em um ambiente com plano de fundo de cor arbitrária.

A partir dessa imagem é aplicada a técnica de *chromakey* em todos os *pixels* da imagem, porém cabe ressaltar que uma modificação na técnica de *chromakey* foi utilizada.

Esta modificação consiste na verificação de três condições para o *chromakey*, ao invés de apenas uma como proposta por (BERGH; LALIOTI, 1999). Assim essa técnica busca limitar se o *pixel* pertence a cor chave pura, ou se o *pixel* pertence a um espectro da cor chave, ou se nao pertence a cor chave. Para melhor exemplificar visualmente essas condições, segue demonstração na Figura 5.1.

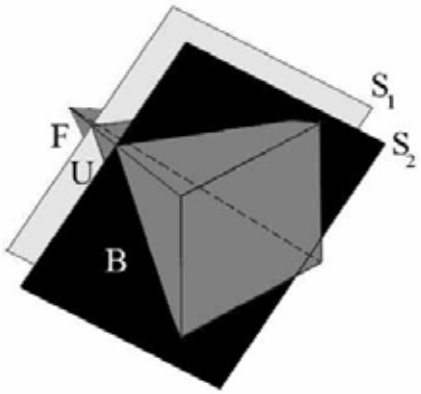


Figura 5.1: *Cubo RGB com a interseção dos planos S1 e S2 (ARRUDA, 2012).*

Após a validação das condições para a técnica de *chromakey* modificada, o resultado obtido é um *trimap* gerado de forma automática sem qualquer interação do usuário.

Com o *trimap* gerado, são estimados os valores de alfa para a região indeterminada do *trimap* com métodos de amostra de cor a partir dos algoritmos *Bayesian matting* (CHUANG et al., 2001) e *Robust matting* (WANG; COHEN, 2007b) ou afinidades da vizinhança dos *pixels* a partir do algoritmo *Geodesic* modificado (ARRUDA, 2012). A partir dos valores estimados de alfa para a região indeterminada, a imagem de entrada juntamente com a imagem do novo plano de fundo, é utilizada a equação da composição (Eq. 2-1) para obter a cor resultante do *pixel* da composição final.

5.4 Considerações Finais

A partir dos métodos utilizando o mapa de profundidade e a técnica de *chromakey*, objetivou-se a geração de *trimap* de forma automática não havendo a necessidade de interações do usuário com marcações na imagem de entrada.

A implementação dos dois métodos propostos é apresentada no próximo capítulo, como etapa necessária para a validação dos mesmos.

CAPÍTULO 6

Implementação dos métodos propostos

Neste capítulo serão apresentadas as implementações dos métodos de geração de *trimap* automático, bem como os materiais utilizados e o ambiente experimental.

6.1 Considerações Iniciais

As implementações descritas nas subseções seguintes envolvem não só os métodos de geração automática de *trimap*, descritos no capitulo anterior, mas todo o *pipeline* de *matting* digital. As soluções foram desenvolvidas em linguagem C#.

6.2 Implementação do método de *trimap* baseado em mapa de profundidade e *chromakey*

A implementação do algoritmo pode ser divida em três fases, sendo melhor exemplificada na Figura 6.1. A primeira fase é representada no módulo captura das imagens pelo *Kinect*, este módulo refere-se à programação utilizando o SDK do *Kinect* para capturar a imagem RGB e o mapa de profundidade e armazená-las para a próxima fase. Neste módulo a primeira atividade a ser realizada é a inicialização do *Kinect*, deixando o dispositivo apto para adquirir os dados dos sensores. Na inicialização do *Kinect* são definidas quais capacidades do *Kinect* serão utilizadas durante a execução do programa. Para a geração do *trimap* automático com uso do *Kinect*, foram utilizadas a câmera RGB, o sensor de profundidade e da informação do corpo humano, para isso, o *Kinect* foi inicializado pela função Initialize do SDK, com os seguintes parâmetros exibidos na Figura 6.2.

Figura 6.1: *Pipeline do funcionamento da geração do trimap a partir do mapa de profundidade.*

```
Runtime nui = new Runtime();
nui.Initialize(RuntimeOptions.UseColor | RuntimeOptions.UseDepthAndPlayerIndex);
nui.VideoStream.Open(ImageStreamType.Video, 2, ImageResolution.Resolution640x480,
                     ImageType.Color);
nui.DepthStream.Open(ImageStreamType.Depth, 2, ImageResolution.Resolution320x240,
                     ImageType.DepthAndPlayerIndex);
```

Figura 6.2: *Trecho de código relativo à inicialização do Kinect.*

Além da inicialização, deve-se abrir o fluxo de imagens RGB e mapa de profundidade com as configurações desejadas, conforme Figura 6.2.

Atualmente, a única configuração de resolução que o *Kinect* permite para a calibração de duas imagens é a resolução de 320x240 para o fluxo de imagens de profundidade e 640x480 para o fluxo de imagens da câmera RGB. Notou-se que estas resoluções são muito baixas para expressar pequenos detalhes nas imagens, sendo esta uma das limitações que estimularam a utilização de uma câmera fullHD em separado.

Após esta fase inicial, procedeu-se a captura dos *frames* dos fluxos da câmera RGB e do mapa de profundidade. Essa captura foi realizada através do Event Model, assim quando um novo quadro de imagem está pronto, um evento é sinalizado e é chamada uma função. A implementação do *Event Model* do SDK pode ser vista na Figura 6.3.

```
void nui_DepthFrameReady(object sender, ImageFrameReadyEventArgs e) {
    PlanarImage planarImage = e.ImageFrame.Image;
    //...
void nui_VideoFrameReady(object sender, ImageFrameReadyEventArgs e){
    PlanarImage planarImage = e.ImageFrame.Image;
    //...
```

Figura 6.3: *Trecho de código que ilustra a leitura dos quadros da câmera RGB e mapa de profundidade do Kinect.*

Após capturar os dois quadros, da câmera RGB e do mapa de profundidade, deve-se fazer a verificação dos corpos humanos reconhecidos pelo dispositivo. Para isso, deve-se percorrer todos os quadros de profundidade e fazer a verificação de cada *pixel* executando-se o código da Figura 6.4. Caso a variável *player* tenha seu valor igual a zero, este *pixel* não pertencerá a nenhum corpo, caso contrário, o número inteiro da variável representará o número do corpo reconhecido pelo *Kinect*.

```
int playerindex = depthFrame16[index] & 7;
if (playerindex == 0) {
    //... pixel não pertence a algum ser humano
} else {
    //... pixel pertence a algum ser humano
}
```

Figura 6.4: *Trecho de código que verifica se o pixel pertence a um ser humano.*

Tendo em vista a diferença de resolução entre a câmera RGB e o mapa de profundidade, é necessário fazer a correspondência do *pixel* no mapa de profundidade ao *pixel* da imagem colorida. Para converter as coordenadas de profundidade para as coordenadas da imagem colorida utiliza-se o método GetColorPixelCoordinatesFromDepthPixel do SDK do *Kinect* transformando o espaço de resolução de 640x480 para 320x240. O trecho do código onde ocorre a calibração entre as imagens pode ser visualizada na Figura 6.5.

```
int index = 0;
int width = planarImage.Width;
int height = planarImage.Height;
for (int y = 0; y < height; y++){
    for (int x = 0; x < width; x++){
        int realDepth = (depthFrame16[index + 1] << 5) | (depthFrame16[index] >> 3);
        int playerindex = depthFrame16[index] & 7;
        byte intensity = (byte)(255 - (255 * realDepth / 0x0fff));
        int colorX;
        int colorY;
        nui.NuiCamera.GetColorPixelCoordinatesFromDepthPixel(resolution, viewArea,
            x, y, (short)(realDepth << 3), out colorX, out colorY);
        colorX = Math.Max(0, Math.Min(nui.VideoStream.Width - 1, colorX));
        colorY = Math.Max(0, Math.Min(nui.VideoStream.Height - 1, colorY));
        int indexColor = (colorY * nui.VideoStream.Width + colorX) * 4;
        //...
    }
}
```

Figura 6.5: *Trecho de código responsável pela calibração dos pixels das imagens RGB e mapa de profundidade.*

No módulo de manipulação dos dados, ocorrem refinamentos na imagem, tendo em vista os ruídos gerados pelo próprio *Kinect* e pela necessidade de fazer a correlação entre a imagem RGB e o mapa de profundidade. Como o *Kinect* é um dispositivo de luz estruturada e o mapa de profundidade em uma visão simplificada é gerado a partir do cálculo da distância entre os pontos da nuvem emitidos pelo projetor infravermelho, é possível perceber que a precisão do mapa de profundidade é proporcional a quantidade de pontos infravermelho emitidos. Assim, o *Kinect* gera ao redor do objeto em primeiro plano, uma área de incerteza contribuindo para a existência de ruídos no mapa de profundidade, juntamente com a aproximação da imagem monocromática para a imagem RGB. A fim de minimizar essa área e os ruídos do mapa de profundidade, utiliza-se a técnica de *chromakey* para refinar a imagem. Portanto, na fase de refinamento há a dependência da comparação do *pixel* do mapa de profundidade com o *pixel* correspondente na imagem RGB:

- Caso o *pixel* no mapa de profundidade represente o ser humano (objeto em primeiro plano) e o seu *pixel* correspondente na imagem RGB não represente a cor do fundo arbitrário (verde ou azul), então esse *pixel* será representado no *trimap* como primeiro plano;
- Caso o *pixel* no mapa de profundidade represente o ser humano (objeto em primeiro plano) e o seu *pixel* correspondente na imagem RGB represente a cor do fundo arbitrário (verde ou azul), então esse *pixel* será representado no *trimap* como área indeterminada;

- Caso o *pixel* no mapa de profundidade não represente o ser humano (objeto em primeiro plano) e o seu *pixel* correspondente na imagem RGB não represente a cor do fundo arbitrário (verde ou azul), então esse *pixel* será representado no *trimap* como área indeterminada;
- Caso o *pixel* no mapa de profundidade não represente o ser humano (objeto em primeiro plano) e o seu *pixel* correspondente na imagem RGB represente a cor do fundo arbitrário (verde ou azul), então esse *pixel* será representado no *trimap* como plano de fundo.

Quando os objetos em primeiro plano são objetos desconhecidos e não são seres humanos, foi utilizada a distância entre o objeto e o *Kinect*, ao invés de utilizar o sistema de detecção de humanos. As demais partes do algoritmo não sofreram modificações para os objetos desconhecidos. O *alphamap* foi obtido através dos algoritmos *Bayesian matting* (CHUANG et al., 2001), *Robust matting* (WANG; COHEN, 2007b) e *Geodesic* Modificado (ARRUDA, 2012). A partir do *alphamap* gerado a composição final é realizada com a execução da equação (Eq. 2-1) para cada *pixel*, desenvolvida por (PORTER; DUFF, 1984).

6.3 Implementação do método de *trimap* baseado em *chromakey*

Com a baixa resolução da imagem RGB e do mapa de profundidade do *Kinect*, foi utilizada uma câmera de alta resolução. Primeiramente foi implementado o algoritmo proposto por (BERGH; LALIOTI, 1999). Este algoritmo consiste em percorrer todos os *pixels* da imagem e verificar o valor da equação (Eq. 2-2) para determinar se esse *pixel* deve ou não ser substituído pelo novo plano de fundo, conforme o código mostrado na Figura 6.6.

```
int tr = 45;
for (int i=0;i < img->height;i++){
    for (int j=0; j < img->width; j++) {
        blue = i * img->widthStep + j * img->nChannels;
        green = i * img->widthStep + j * img->nChannels + 1;
        red = i * img->widthStep + j * img->nChannels + 2;
        //formula simplificada
        float formula = (2 * data[green] -  data[blue] - data[red]);
        //formula completa
        // formula = (float) sqrt((double)(data[green] - data[red]) *
        //            (data[green] - data[red]) + (data[green] - data[blue])
        //                * (data[green] - data[blue]));
        if (formula < tr){
            // pixel pertence ao fundo verde
            //...
        } else {
            // pixel não pertence ao fundo verde
            //...
        }
    }
}
```

Figura 6.6: *Trecho de código que ilustra a implementação do chromakey com base no método de (BERGH; LALIOTI, 1999).*

Com o desenvolvimento do algoritmo de *chromakey*, foi possível gerar uma nova imagem de saída, a qual originou o *trimap* automático. A diferença entre os dois códigos está nas três condições, ao invés de uma, como no caso anterior.

Conforme descrito na subseção 5.3, a primeira condição ocorre para limitar os *pixels* marcados como plano da frente, a segunda para demarcar os *pixels* como plano de fundo, aos demais *pixels* foram marcados como pontos indeterminados. Todos eles utilizando valores da Equação (Eq. 2-2). Essa implementação pode ser visualizada na Figura 6.7.

```
int tr = 45;
int tf = 190;
for ( int i=0;i < img->height;i++){
    for (int j=0; j < img->width; j++) {
        blue = i * img->widthStep + j * img->nChannels;
        green = i * img->widthStep + j * img->nChannels + 1;
        red = i * img->widthStep + j * img->nChannels + 2;
        float formula = (2 * data[green] -  data[blue] - data[red]);
        //formula = (float) sqrt((double)(data[green] - data[red]) *
        //(data[green] - data[red]) + (data[green] - data[blue]) * (data[green] - data[blue]));
        if (formula < tr){ // se é verde
            data4[blue]=255;data4[green]=255;data4[red]=255;//branco
        } else {
            if (formula < tf){ //se estiver entre as faixas de verde
                data4[blue]= 125; data4[green]= 125;data4[red]= 125;//cinza
            } else { // é plano de fundo
                data4[blue]=0; data4[green]=0; data4[red]=0;//preto
            }
        }
    }
}
```

Figura 6.7: *Trecho de código que ilustra a implementação do chromakey modificado para geração do trimap automático.*

Após a geração do *trimap* automático, o *alphamap* é gerado a partir através dos algoritmos *Bayesian matting* (CHUANG et al., 2001), *Robust matting* (WANG; COHEN, 2007b) e *Geodesic* Modificado (ARRUDA, 2012). Com o *alphamap* gerado, a composição final é executada por meio da equação (Eq. 2-1) para cada *pixel*, criada por (PORTER; DUFF, 1984).

6.4 Considerações finais

Com a implementação dos métodos propostos construiu-se o *pipeline* do *matting* digital completo. No próximo capítulo serão apresentados e analisados os testes realizados com as implementações desenvolvidas e descritas no presente capítulo.

CAPÍTULO 7

Experimentos e Análise de Resultados

Neste capítulo serão mostrados os testes realizados, juntamente com suas análises e resultados.

7.1 Considerações Iniciais

Foram realizados testes com a finalidade de comparar-se, usando-se métricas quantitativas, as imagens produzidas utilizando *trimaps* gerados de forma automática e *trimaps* produzidos manualmente. Para a realização dos testes foram utilizados os algoritmos de *matting* digital, o *bayesian, geodesic* modificado, e o *Robust matting*.

O ambiente experimental constitui-se de um ambiente composto de um plano de fundo com de 6,5 metros de comprimento por 2,5 metros pintada na cor verde, destinada à aplicação da técnica de *chromakey*, conforme mostra a Figura 7.1.

Os experimentos foram realizados com luz controlada, onde foram utilizadas duas luminárias com uma lâmpada fria de 20W e 3 lâmpadas tipo T3 de 20W devidamente posicionadas conforme as Figura 7.1(a) e 7.1(b).

(a)

(b)

Figura 7.1: *Ambiente experimental.*

Além dessa estrutura física, foram utilizados um notebook com processador Core i5, 4Gb de memória, disco rígido de 640Gb, sistema operacional Windows Seven, juntamente com o *Kinect* da Microsoft e a câmera fullHD Logitech modelo C910.

7.2 Métricas utilizadas

Para interpretação dos resultados foram utilizadas as métricas descritas nessa subseção.

7.2.1 *Strutural Similarity* - SSIM

O Índice de Similaridade Estrutural (*Strutural Similarity* - SSIM) foi proposto (WANG et al., 2004) com o objetivo de mensurar as correlações locais em luminância, contraste e estrutura entre duas imagens. Para o cálculo do índice é comparada uma imagem sem distorção com a imagem que à ser analisada. Esta metodologia é uma tentativa de analisar a imagem de acordo com o olho humano, fazendo análises na mudança estrutural das imagens. O cálculo é feito pela Equação (Eq. 7-1), em cada janela.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{7-1}$$

Onde:

x representa a janela da primeira imagem;
y representa a janela da segunda imagem;
μ_x é a média de x;
μ_y é a média de y;
σ_x^2 é a variância de x;
σ_y^2 é a variância de y;
σ_{xy} é a covariância entre x e y;
$C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$;
L é o valor máximo dos pixeis (255 para imagens em escala de cinza), C_1 e C_2 são constantes definidas por (WANG et al., 2004) como 0,01 e 0,03 respectivamente;
k_1 e k_2 são constantes que devem ter valores muito pequenos.

Para medir a qualidade da imagem em sua totalidade, é utilizada a média dos valores calculados em todas as janelas, sendo que é criada uma para cada *pixel*.

O resultado pode variar entre -1 e 1, quanto mais próximo de 1 for o valor do índice SSIM, melhor será a qualidade estrutural da imagem, ou seja, as imagens são mais parecidas.

7.2.2 *Universal Image Quality Index* - UIQI

O índice UIQI (*Universal Image Quality Index*) foi proposto por (ZHOU; BOVIK, 2002) para medir a similaridade entre duas imagens. O UIQI foi projetado para modelar qualquer distorção na imagem como uma combinação de três fatores: perda de correlação, distorção luminância e distorção no contraste. O UIQI é definido pela equação (Eq. 7-2).

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x}^2 + \bar{y}^2)]} \tag{7-2}$$

Onde:

x representa os *pixels* da primeira imagem;
y representa os *pixels* da segunda imagem;
$\bar{x}$ é o valor médio de x;
$\bar{y}$ é o valor médio de y;
σ_x^2 é a variância de x;
σ_y^2 é a variância de y;
σ_{xy} é a covariância entre x e y;

O valor de Q pode variar entre -1 e 1, sendo 1 para as imagens totalmente iguais e -1 com as imagens extremamente diferentes.

7.3 Resultados utilizando o *trimap* gerado a partir do mapa de profundidade

Foram realizados testes utilizando a automatização do *trimap* a partir do *Kinect* juntamente com a técnica de *chromakey*. A Figura 7.2 mostra a imagem original da qual foi extraído o objeto em primeiro plano, a Figura 7.3 mostra o mapa de profundidade e a Figura 7.4 mostra o *trimap* gerado de forma automática pelo método.

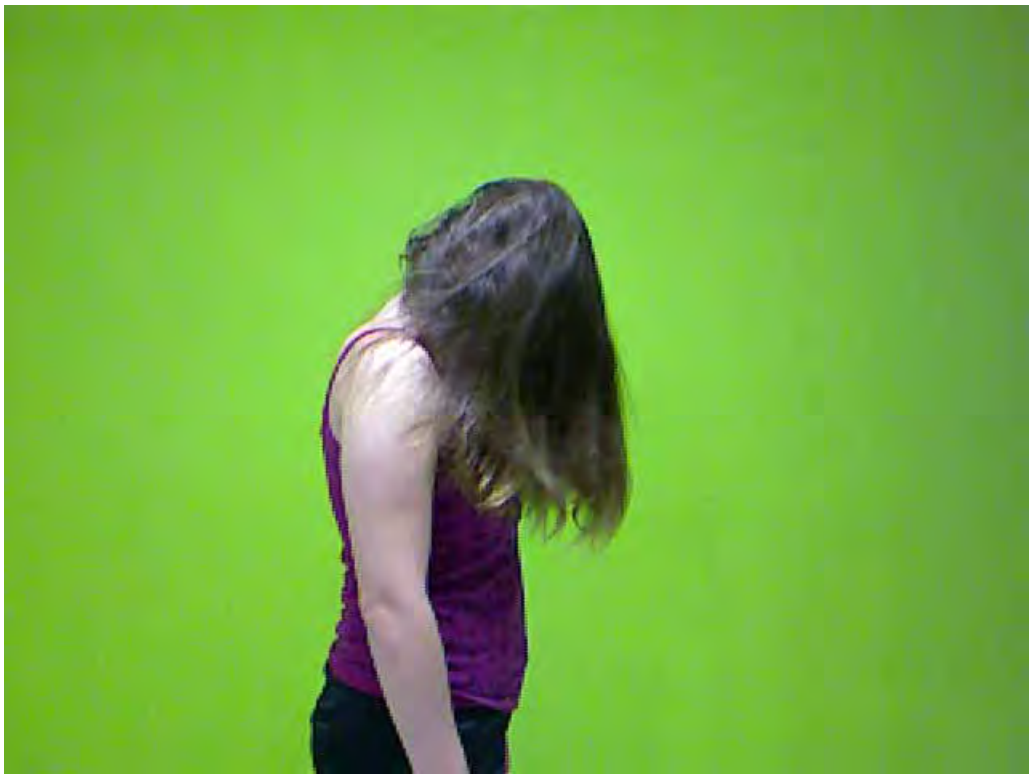


Figura 7.2: *Imagem de entrada capturada pela câmera RGB do Kinect com fundo verde.*

Figura 7.3: *Mapa de profundidade gerado pelo Kinect.*

Figura 7.4: *Trimap gerado pelo método desenvolvido.*

A Figura 7.5 mostra o *matting* digital realizado com o algoritmo *Bayesian matting* (CHUANG et al., 2001). Na Figura 7.6 tem-se o *matting* digital produzido pelo algoritmo *Robust matting* (WANG; COHEN, 2007b). A Figura 7.7 mostra o *matting* a partir do método *Geodesic* modificado (ARRUDA, 2012). Com objetivo de verificar a eficiência do *trimap* gerado de forma automática nos algoritmos de *matting* digital, foi utilizado um fundo branco como novo plano de fundo.

Figura 7.5: *Recorte gerado pelo Bayesian matting (CHUANG et al., 2001), com o trimap automático.*

Figura 7.6: *Recorte gerado pelo Robust matting (WANG; COHEN, 2007b), com o trimap automático.*

Figura 7.7: *Recorte gerado pelo Geodesic modificado (ARRUDA, 2012), com o trimap automático.*

Para cada teste realizado criou-se um *trimap* pelo método manual, para fins de comparação. Assim a Figura 7.8 exibe o *trimap* gerado de forma manual. A Figura 7.9 ilustra o *matting* digital realizado com o algoritmo *Bayesian matting* (CHUANG et al., 2001). A Figura 7.10 apresenta o *matting* digital produzido pelo algoritmo *Robust matting* (WANG; COHEN, 2007b), e a Figura 7.11 mostra o *matting* a partir do método *Geodesic* modificado (ARRUDA, 2012).

Figura 7.8: *Trimap gerado de forma manual.*

Figura 7.9: *Recorte gerado pelo Bayesian matting (CHUANG et al., 2001), com o trimap manual.*

Figura 7.10: *Recorte gerado pelo Robust matting (WANG; COHEN, 2007b), com o trimap manual.*

Figura 7.11: *Recorte gerado pelo Geodesic modificado (ARRUDA, 2012), com o trimap manual.*

Semelhante aos testes anteriores, foram realizados os mesmos testes, porém com um novo plano de fundo. A partir destes testes foi possível verificar como as falhas dos métodos são menos perceptíveis, como ilustrados nas Figuras 7.21, 7.22 e 7.23.

Figura 7.12: *Composição gerado pelo Bayesian matting (CHUANG et al., 2001), com o novo plano de fundo.*

Figura 7.13: *Composição gerado pelo Robust matting (WANG; COHEN, 2007b), com o novo plano de fundo.*

Figura 7.14: *Composição gerado pelo Geodesic modificado (ARRUDA, 2012), com o novo plano de fundo.*

Com a finalidade de verificar-se o desempenho na geração do *trimap* automático, mensurou-se o tempo de execução dos métodos de *matting* implementados. Conforme mostra a Tabela 7.1, bem como o *trimap* automático gerado e o *trimap* manual.

Tabela 7.1: *Tempo de processamento dos métodos de matting digital em segundos, a partir do Kinect.*

Imagens[1]	Bayesian		Robust		Geodesic modificado	
	Trimap Manual	Trimap Automático	Trimap Manual	Trimap Automático	Trimap Manual	Trimap Automático
Img_1	8,84	9,12	38,19	66,01	0,38	0,40
Img_2	6,32	6,14	22,36	18,85	0,24	0,25
Img_3	7,65	5,89	42,25	31,52	0,21	0,19
Img_4	10,87	5,51	48,86	16,04	0,27	0,18
Img_5	9,62	5,90	37,25	18,02	0,24	0,19

[1]As imagens referenciadas nesta tabela encontram-se no Apêndice A.

A Tabela 7.2 mostra, em porcentagem, o tempo de processamento do *trimap* automático em relação ao *trimap* manual.

Tabela 7.2: *Trimap Automático / Trimap Manual (%)*

Imagens[1]	Bayesian	Robust	Geodesic modificado
Img_1	1,0316%	1,7284%	1,0526%
Img_2	0,9715%	0,8430%	1,0416%
Img_3	0,7699%	0,7460%	0,9047%
Img_4	0,5068%	0,3282%	0,6666%
Img_5	0,6133%	0,4837%	0,7916%

[1]As imagens referenciadas nesta tabela encontram-se no Apêndice A.

A Tabela 7.3 ilustra a diferença entre o tempo de processamento do *trimap* manual e o *trimap* automático em segundos.

Tabela 7.3: *Trimap Manual - Trimap Automático (em segundos).*

Imagens[1]	Bayesian	Robust	Geodesic modificado
Img_1	-0,28	-27,82	-0,02
Img_2	0,18	3,51	-0,01
Img_3	1,76	10,73	0,02
Img_4	5,36	32,82	0,09
Img_5	3,72	19,23	0,05

[1]As imagens referenciadas nesta tabela encontram-se no Apêndice A.

Após a realização dos testes descritos, foi possível comprovar que a baixa resolução das imagens obtidas por meio da câmera RGB do *Kinect*, contribuiram para o aumento da área indeterminada do *trimap*. Contudo foi possível perceber também que os resultados obtidos com o método *Bayesian matting* foram mais precisos do que os obtidos com o uso do *Robust matting*. Esses motivos explicam o tempo de execução do *trimap* automático ser maior que o *trimap* manual para alguns exemplos, como demonstrado na Tabela 7.1.

7.3.1 Análise da qualidade do *trimap* automático

Nesta subseção são analisados os *trimaps* gerados de forma automática a partir da imagem de profundidade do *Kinect* e a técnica de *chromakey*.

Para analisar a eficiência do *trimap* automático em relação ao *trimap* manual, foram utilizados os testes UIQI e SSIM para todas as imagens geradas pelos métodos de *matting* digital capturadas a partir do *Kinect*. A Tabela 7.4 exibe os resultados obtidos.

Tabela 7.4: *Comparação de qualidade entre as imagens geradas pelos métodos de matting digital com trimap automático e trimap manual, a partir do Kinect.*

Imagens[1]	Bayesian		Robust		Geodesic modificado	
	UIQI	SSIM	UIQI	SSIM	UIQI	SSIM
Img_1	0,9999	0,9988	0,9997	0,9961	0,9994	0,9933
Img_2	0,9994	0,9963	0,9984	0,9914	0,9992	0,9933
Img_3	0,9997	0,9976	0,9998	0,9984	0,9998	0,9984
Img_4	0,9995	0,9961	0,9994	0,9954	0,9997	0,996
Img_5	0,9998	0,9978	0,9995	0,9958	0,9998	0,998

[1] As imagens referenciadas nesta tabela encontram-se no Apêndice A.

Com estes resultados é possível perceber que as composições finais ficaram muito próximas, mesmo geradas a partir de *trimaps* diferentes, pois para todos os testes realizados os valores se aproximaram de 1 das imagens resultantes.

7.4 Resultados utilizando o *trimap* gerado por *chromakey*

Para os próximos testes, a captura da imagem foi obtida a partir de uma câmera fullHD, para este ambiente a câmera foi acoplada na parte superior do *Kinect*, conforme Figura 7.15.

Figura 7.15: *Câmera fullHD acoplada ao Kinect.*

Os testes continuam utilizando o ambiente de *chromakey* com fundo verde, conforme mostrado na Figura 7.16. Os testes foram realizados com o ator a uma distância de 90cm da câmera e a 100cm do plano de fundo, gerando o *trimap* de forma automática da Figura 7.17.

Figura 7.16: *Imagem capturada pela câmera fullHD.*

Figura 7.17: *Trimap automático gerado a partir da imagem RGB da câmera fullHD.*

A partir do *trimap* gerado, é possível observar que a área indeterminada (área em cinza) contém poucos *pixels*, diminuindo-se, com isso, diminui a área de cálculo do alfa. Os testes realizados utilizaram as implementações dos métodos *Bayesian matting* (CHUANG et al., 2001), *Robust matting* (WANG; COHEN, 2007b) e o *Geodesic* modificado (ARRUDA, 2012). Com a finalidade de demonstrar a eficiência do *trimap* gerado de forma automática nos algoritmos de *matting* digital, foi utilizado um fundo branco como novo plano de fundo, conforme mostram as Figuras 7.18, 7.19 e 7.20.

Figura 7.18: *Recorte gerado pelo Bayesian matting (CHUANG et al., 2001), com o trimap automático.*

Figura 7.19: *Recorte gerado pelo Robust matting (WANG; COHEN, 2007b), com o trimap automático.*

Figura 7.20: *Recorte gerado pelo Geodesic modificado (ARRUDA, 2012), com o trimap manual.*

As Figuras 7.21, 7.22 e 7.23 ilustram testes semelhante aos anteriores, porém utilizando-se uma nova imagem de fundo.

Figura 7.21: *Composição gerado pelo Bayesian matting (CHUANG et al., 2001), com o novo plano de fundo.*

Figura 7.22: *Composição gerado pelo Robust matting (WANG; COHEN, 2007b), com o novo plano de fundo.*

Figura 7.23: *Composição gerado pelo Geodesic modificado (ARRUDA, 2012), com o novo plano de fundo.*

Outro teste realizado com a finalidade de observar a eficiência do *trimap* automático, foi mensurar o tempo de execução dos métodos de *matting* implementados. Conforme mostra a Tabela 7.5.

Tabela 7.5: *Tempo de processamento dos métodos de matting digital em segundos, a partir do Chromakey.*

Imagens[1]	Bayesian		Robust		Geodesic modificado	
	Trimap Manual	Trimap Automático	Trimap Manual	Trimap Automático	Trimap Manual	Trimap Automático
Img_6	8.640,19	692,51	45.047,93	3.324,78	115,17	29,86
Img_7	38.741,98	4.170,49	109.720,55	35.117,07	146,02	118,39
Img_8	6.129,16	118,30	69.828,67	3.495,56	116,35	30,71
Img_9	4.382,08	107,80	58.885,45	3.210,50	110,29	27,88

[1] As imagens referenciadas nesta tabela encontram-se no Apêndice A.

Após a realização dos testes descritos, foi possível observar que quando o *trimap* possui uma alta proporção de *pixels* indeterminados, o Baeysian *matting* gera resultados

visualmente melhores do que no *Robust matting*. No entanto quando utilizado *trimaps* com baixa proporção de *pixels* indeterminados, os resultados do *Robust matting* são visualmente superiores, como ilustrados nas Figuras 7.18 e 7.19.

A Tabela 7.6 mostra, em porcentagem, o tempo de processamento do *trimap* automático em relação ao *trimap* manual.

Tabela 7.6: *Trimap Automático / Trimap Manual (%)*

Imagens[1]	Bayesian	Robust	Geodesic modificado
Img_6	8%	7%	26%
Img_7	11%	32%	81%
Img_8	2%	5%	26%
Img_9	2%	5%	25%

[1]As imagens referenciadas nesta tabela encontram-se no Apêndice A.

A Tabela 7.7 ilustra a diferença entre o tempo de processamento do *trimap* manual e o *trimap* automático em segundos.

Tabela 7.7: *Trimap Manual - Trimap Automático (em segundos).*

Imagens[1]	Bayesian	Robust	Geodesic modificado
Img_6	7.947,68	41.723,15	85,31
Img_7	34.571,49	74.603,48	27,63
Img_8	6.010,86	66.333,11	85,64
Img_9	4.274,28	5.5674,95	82,41

[1]As imagens referenciadas nesta tabela encontram-se no Apêndice A.

7.4.1 Análise da qualidade do *trimap* automático

Nesta subseção são analisados os *trimap* gerados de forma automática por *chromakey* com a câmera fullHD.

Para analisar a eficiência do *trimap* automático em relação ao *trimap* manual, foram utilizados os testes UIQI e SSIM para todas as imagens geradas pelos métodos de *matting* digital capturadas a partir do *Kinect*. Desta forma, foram comparadas as imagens das Figuras 7.18, 7.19, 7.20, juntamente com as imagens geradas a partir do *trimap* manual, conforme ilustra a Tabela 7.8.

Tabela 7.8: *Comparação de qualidade entre as imagens geradas pelos métodos de matting digital com trimap automático e trimap manual.*

Imagens[1]	Bayesian		Robust		Geodesic modificado	
	UIQI	SSIM	UIQI	SSIM	UIQI	SSIM
Img_6	0,9851	0,9628	0,9897	0,9893	0,9963	0,9879
Img_7	0,9738	0,9189	0,9858	0,9786	0,9844	0,9718
Img_8	0,9879	0,9706	0,9885	0,9883	0,9902	0,9746
Img_9	0,9892	0,9786	0,9911	0,9927	0,9934	0,979

[1]As imagens referenciadas nesta tabela encontram-se no Apêndice A.

Com os resultados obtidos foi possível reconhecer que as composições finais ficaram semelhantes, mesmo sendo geradas a partir de *trimaps* diferentes, uma vez que todos os testes realizados os valores se aproximaram de 1, para os métodos de *matting* utilizados.

7.5 Testes em vídeo

Os testes em vídeo foram realizados com a câmera fullHD, e o *trimap* foi gerado automaticamente com auxílio do *chromakey*. Os testes foram realizados com uma resolução de 640x480 a uma taxa de 15 FPS(*Frames per second*), ou seja, a taxa de quadros por segundo na qual o vídeo é exibido. Os testes foram realizados com o método do *Geodesic* modificado (ARRUDA, 2012), tendo em vista os resultados obtidos nos testes anteriores sua agilidade em comparação com o *Bayesian* e o *Robust Matting*.

O resultado pode ser observado na Figura 7.24, onde a composição final da imagem foi combinada com um fundo na cor vermelha.

Figura 7.24: *Representação de um dos quadros do vídeo após o matting digital com trimap automático.*

Apesar de terem sido realizados poucos experimentos envolvendo vídeos, os métodos desenvolvidos demonstraram possuir potencial para isto. No entanto, a comprovação de sua eficiência requer uma análise mais profunda e estão fora do escopo delineado para o presente trabalho.

7.6 Considerações finais

O *trimap* gerado automaticamente a partir do *Kinect* ou da câmera fullHD, tendo em vista o alto grau de similaridade entre os resultados obtidos com *trimap* automático e *trimap* manual, demonstrou a eficácia dos métodos de geração automática de *trimap* para imagem. Notou-se que, potencialmente, os métodos implementados poderiam ser aplicados no processo de *matting* digital envolvendo vídeos, mas isto demanda maiores investigações.

CAPÍTULO 8

Conclusões e Trabalhos Futuros

Neste capítulo serão apresentadas as conclusões e sugestões para trabalhos futuros.

8.1 Conclusões

Na comparação entre os resultados obtidos com o *trimap* automático e o *trimap* manual, o primeiro mostrou-se eficiente tendo em vista o alto grau de similaridade das imagens resultantes para os métodos de *matting* digital testados. Outro fator relevante é a não interação do usuário com o método, não havendo a necessidade de qualquer marcação manual na imagem de entrada.

O *trimap* automático é uma informação extra a qual pode ser agregada em outros métodos de *matting* digital tanto para imagem como vídeo, ampliando suas possibilidades de utilização.

A utilização do *Kinect* como informação extra possibilita a geração do *trimap* em ambientes mais simples, onde a resolução da imagem não seja determinante no processo. Porém, o método de geração do *trimap* é eficaz e com o avanço a resolução deste dispositivo poderá vir a produzir resultados mais expressivos.

A utilização da câmera fullHD juntamente com a técnica de *chromakey* na geração do *trimap* automático, demonstrou resultados interessantes tanto para imagem como vídeo. No entanto vale lembrar que a marcação manual de *trimap* em vídeos é muito custosa, tendo em vista a necessidade da escolha dos quadros chaves além do volume de quadros por segundo.

Assim, o *trimap* automático é mais uma ferramenta auxiliar para as soluções de *matting* digital tanto para imagem como vídeo.

Apesar da similaridade entre os resultados obtidos pelos dois métodos (manual

e automático), as composições resultantes apresentam *aliasing* (serrilhamento), o que indica a necessidade adicional da aplicação de técnicas de *anti-aliasing*. A aplicação de tais técnicas, no entanto, estão fora do escopo determinado para o presente trabalho.

8.2 Trabalhos Futuros

Como desdobramento deste trabalho, sugere-se os seguintes trabalhos futuros:

- Validar e aprimorar o algoritmo de geração de *trimap* automático para objetos transparentes, ou quando há fumaça em cena.

- Verificar a possibilidade de utilização do algoritmo de *trimap* automático em vídeo de alta resolução em tempo real.

- Uso do *Kinect* como informação extra, acoplado juntamente com uma câmera fullHD, utilizando ou não a técnica de *chromakey*. Isto viabilizaria sua utilização em aplicações de Estúdios Virtuais e Videoconferência.

- Utilização de câmera *time of flight* como informação extra, utilizando ou não a técnica de *chromakey*.

- Utilização de câmera *time of flight* acoplada com o *Kinect* e com uma câmera fullHD, a fim de melhor a presição do mapa de profundidade.

- Estudo e aplicação de técnicas de *anti-alising*.

- Utilização do *Kinect for Windows* que permite a aproximação do objeto até 40cm do dispositivo, ao invés de 80 cm do *Kinect* do Xbox.

Referências Bibliográficas

ARRUDA, H. F. **Paralelização de algoritmos de trimap utilizando a arquitetura CUDA**. Monografia (Trabalho de conclusão) — Universidade Estadual Paulista Julio de Mesquita Filho, Bauru - SP, 2012.

BAI, X.; SAPIRO, G. Geodesic matting: A framework for fast interactive image and video segmentation and matting. **Int. J. Comput. Vision**, Kluwer Academic Publishers, Hingham, MA, USA, v. 82, n. 2, p. 113–132, abr. 2009. ISSN 0920-5691. Disponível em: <http://dx.doi.org/10.1007/s11263-008-0191-z>.

BERGH, V. D.; LALIOTI, V. Software chroma keying in an immersive virtual environment. **South African Computer Journal**, v. 24, p. 155–162, 1999.

BIEDNY, B. M. D.; MOODY, N. **Photoshop Channel Chops**. [S.l.]: New Riders Publishing, 1998.

BRINKMANN, R. **The art and science of digital compositing**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999. ISBN 0-12-133960-2.

CANNY, J. A computational approach to edge detection. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 8, p. 679–698, November 1986. ISSN 0162-8828. Disponível em: <http://dl.acm.org/citation.cfm?id=11274.11275>.

CHUANG, Y.-Y. **New models and methods for matting and compositing**. Tese (Doutorado) — University of Washington, Washington - USA, 2004.

CHUANG, Y.-Y. et al. A bayesian approach to digital matting. In: **Proceedings of 2001 International Conference on Computer Vision and Pattern Recognition**. Kauai, HI, USA: [s.n.], 2001. II, p. 264–271.

FOSTER, J. **The Green Screen Handbook: Real-World Production Techniques**. [S.l.]: Sybex, 2010.

GVILI, R. et al. Depth keying. In: WOODS, A. J. et al. (Ed.). Santa Clara, CA, USA: SPIE, 2003. v. 5006, n. 1, p. 564–574.

HU, Y. Z. H. Robust matting based on region growing. In: . [S.l.: s.n.], 2010.

ISAKSEN, A.; MCMILLAN, L.; GORTLER, S. J. Dynamically reparameterized light fields. In: **Proceedings of the 27th annual conference on Computer graphics and interactive techniques**. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000. (SIGGRAPH '00), p. 297–306. ISBN 1-58113-208-5.

JOSHI, N.; MATUSIK, W.; AVIDAN, S. Natural video matting using camera arrays. In: **ACM SIGGRAPH 2006 Papers**. New York, NY, USA: ACM, 2006. (SIGGRAPH '06), p. 779–786.

KUHN, T. **The Kinect Sensor Platform**. 2011. Disponível em: <http://www.lmt.ei.tum.de-/courses/hsmt/proceedings/aimt_ss11.pdf>.

LEVIN, A.; RAV-ACHA, A.; LISCHINSKI, D. Spectral matting. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 30, p. 1699–1712, October 2008. ISSN 0162-8828.

LIMBERGER, F. A. **Geração de Trimap em tempo real utilizando o Kinect**. Monografia (Trabalho de conclusão) — Universidade Federal de Santa Maria, Santa Maria - RS, 2011.

MCGUIRE, M.; MATUSIK, W.; YERAZUNIS, W. Practical, Real-time Studio Matting using Dual Imagers. In: . [s.n.], 2006. p. 235–244. Disponível em: <http://www.eg.org/EG/DL-/WS/EGWR/EGSR06/235-244.pdf>.

MONACO, J. **How to Read a Film: The World of Movies, Media, Multimedia:Language, History, Theory**. 3rd. ed. [S.l.]: Oxford University Press, 2000.

MONTABONE, S. **Beginning Digital Image ProcessingUsing Free Tools for Photographers**. [S.l.]: Apress, 2010.

PORTER, T.; DUFF, T. Compositing digital images. In: **Proceedings of the 11th annual conference on Computer graphics and interactive techniques**. New York, NY, USA: ACM, 1984. (SIGGRAPH '84), p. 253–259. ISBN 0-89791-138-5.

RHEMANN, C. et al. High resolution matting via interactive trimap segmentation. In: **CVPR'08**. [S.l.: s.n.], 2008.

RICKITT, R. **Special Effects: The History and Technique**. [S.l.]: Watson-Guptill Publications, 2000. Hardcover. ISBN 0823077330.

RUZON, M. A.; TOMASI, C. **Alpha Estimation in Natural Images**. 2000.

SHOTTON, J. et al. Real-Time Human Pose Recognition in Parts from Single Depth Images. In: . [s.n.], 2011. Disponível em: <http://research.microsoft.com/apps/pubs-/default.aspx?id=145347>.

SMITH, A. R.; BLINN, J. F. Blue screen matting. In: **Proceedings of the 23rd annual conference on Computer graphics andinteractive techniques**. New York, NY, USA: ACM, 1996. (SIGGRAPH '96), p. 259–268. ISBN 0-89791-746-4. Disponível em: <http://doi.acm.org/10.1145/237170.237263>.

SUN, J. et al. Flash matting. In: **ACM SIGGRAPH 2006 Papers**. New York, NY, USA: ACM, 2006. (SIGGRAPH '06), p. 772–778. ISBN 1-59593-364-6.

WANG, J.; COHEN, M. F. Image and video matting: a survey. **Found. Trends. Comput. Graph. Vis.**, Now Publishers Inc., Hanover, MA, USA, v. 3, p. 97–175, January 2007. ISSN 1572-2740. Disponível em: <http://dl.acm.org/citation.cfm?id=1391083.1391084>.

WANG, J.; COHEN, M. F. Optimized color sampling for robust matting. In: **2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), 18-23 June 2007, Minneapolis, Minnesota, USA**. Minnesota, USA: IEEE Computer Society, 2007.

WANG, L. et al. **TofCut: Towards Robust Real-time Foreground Extraction Using a Time-of-Flight Camera**. 2010.

WANG, Z. et al. Image quality assessment: From error visibility to structural similarity. **IEEE TRANSACTIONS ON IMAGE PROCESSING**, v. 13, n. 4, p. 600–612, 2004.

ZHOU, W.; BOVIK, A. C. A Universal Image Quality Index. **IEEE SIGNAL PROCESSING LETTERS**, XX, mar. 2002.

APÊNDICE A

Imagens dos testes

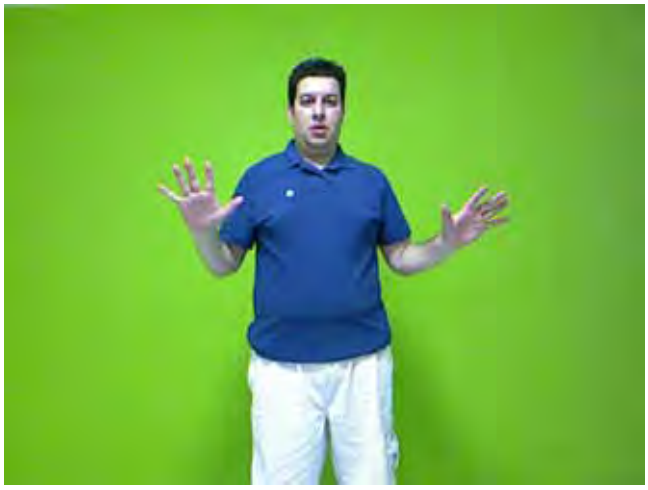


Figura A.1: *Img_1 - Imagem capturada pela câmera RGB*

Figura A.2: *Img_1 - Trimap manual*

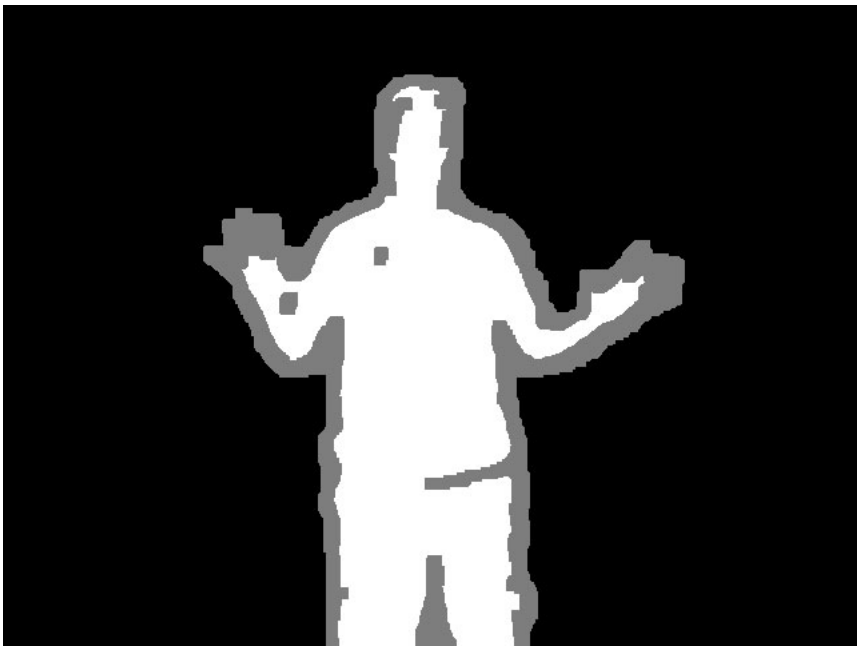


Figura A.3: *Img_1 - Trimap automático*

Figura A.4: *Img_1 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.5: *Img_1 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.6: *Img_1 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.7: *Img_1 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.8: *Img_1 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.9: *Img_1 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

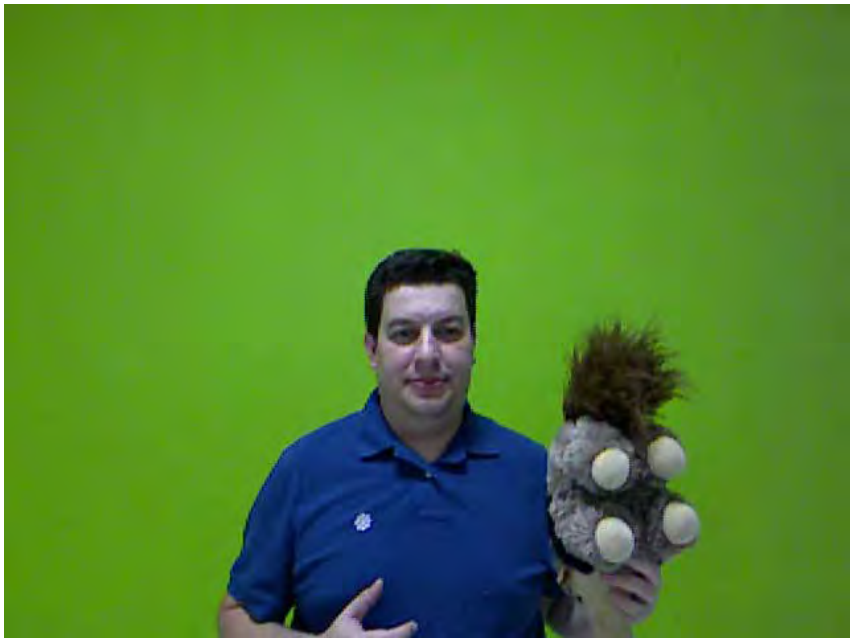


Figura A.10: *Img_2 - Imagem capturada pela câmera RGB*

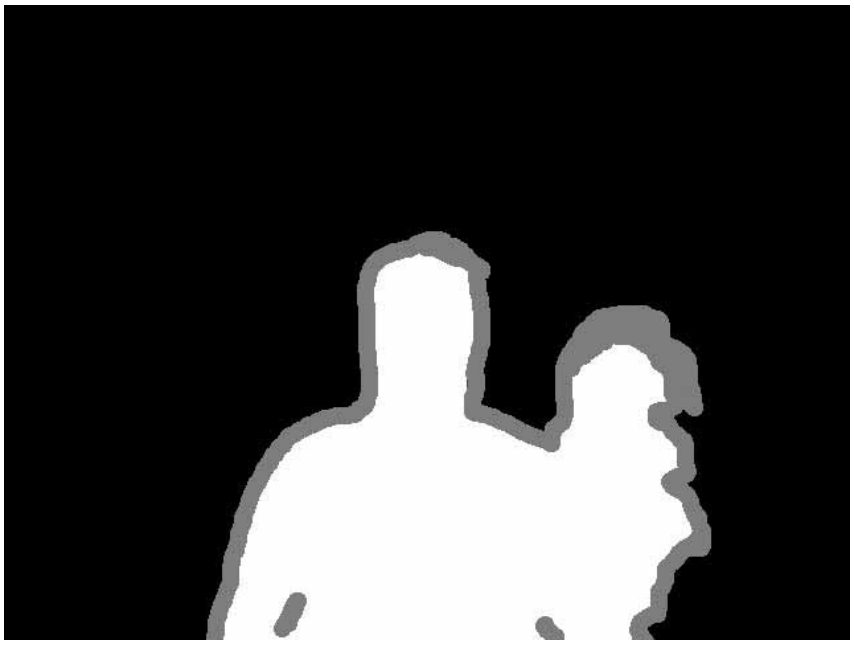


Figura A.11: *Img_2 - Trimap manual*

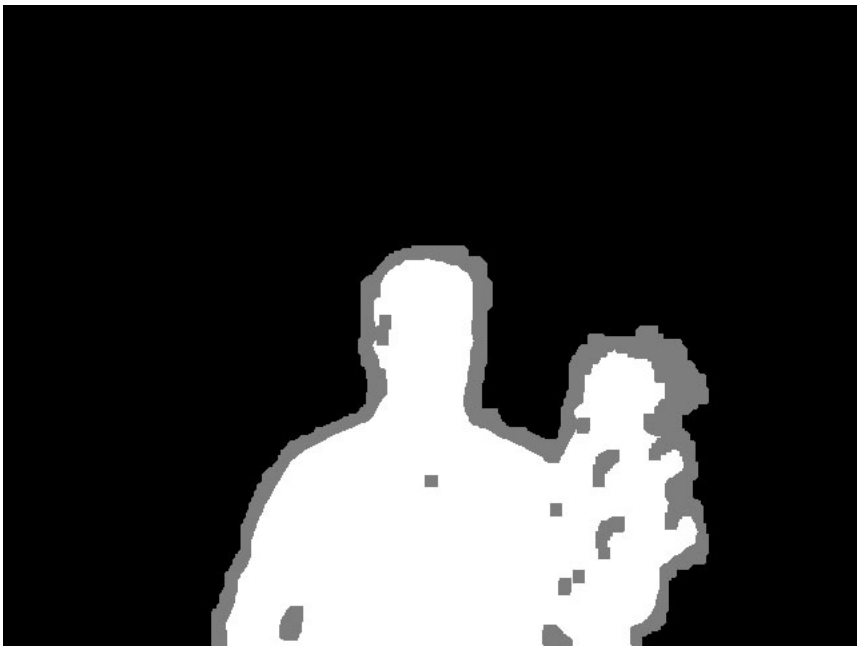


Figura A.12: *Img_2 - Trimap automático*

Figura A.13: *Img_2 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.14: *Img_2 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.15: *Img_2 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.16: *Img_2 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.17: *Img_2 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.18: *Img_2 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

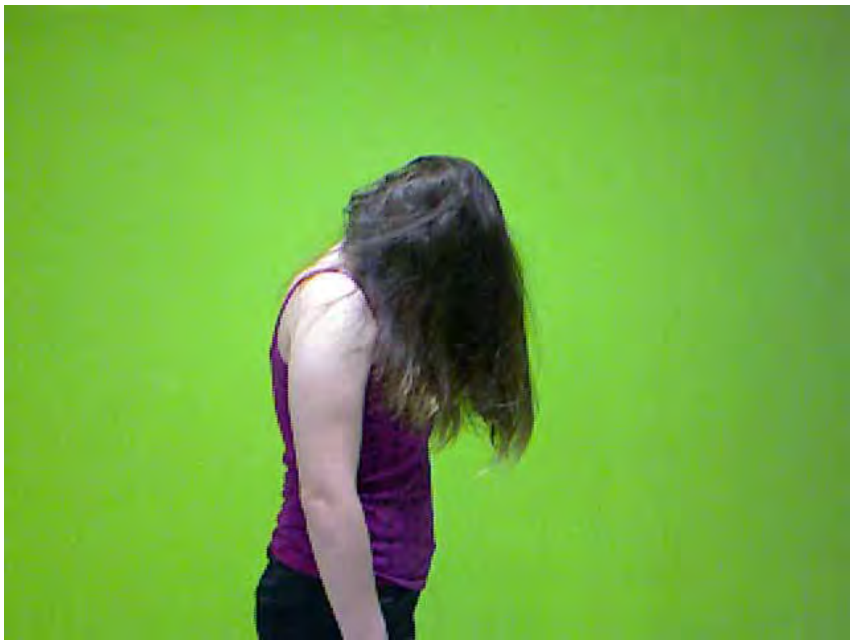


Figura A.19: *Img_3 - Imagem capturada pela câmera RGB*

Figura A.20: *Img_3 - Trimap manual*

Figura A.21: *Img_3 - Trimap automático*

Figura A.22: *Img_3 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.23: *Img_3 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.24: *Img_3 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.25: *Img_3 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.26: *Img_3 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.27: *Img_3 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

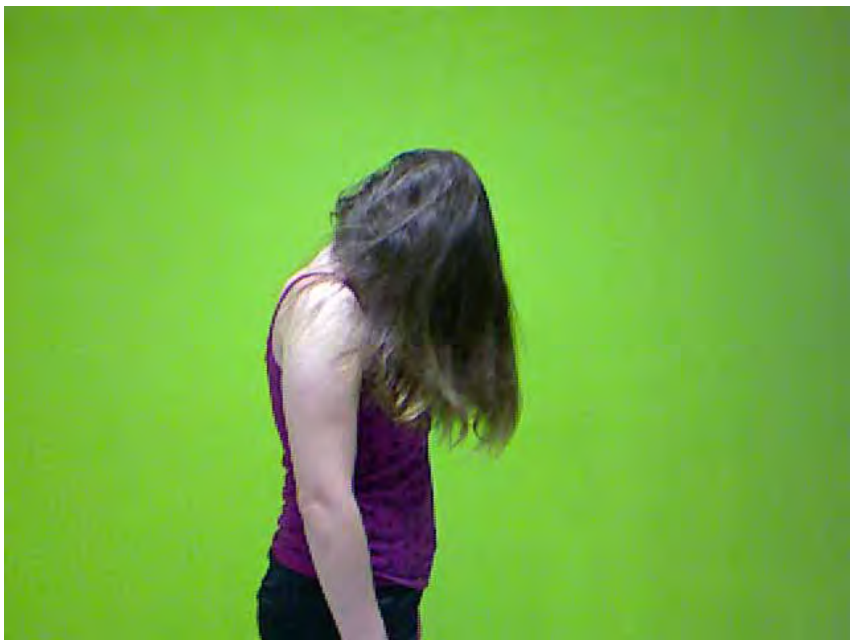


Figura A.28: *Img_4 - Imagem capturada pela câmera RGB*

Figura A.29: *Img_4 - Trimap manual*

Figura A.30: *Img_4 - Trimap automático*

Figura A.31: *Img_4 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.32: *Img_4 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.33: *Img_4 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.34: *Img_4 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.35: *Img_4 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.36: *Img_4 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

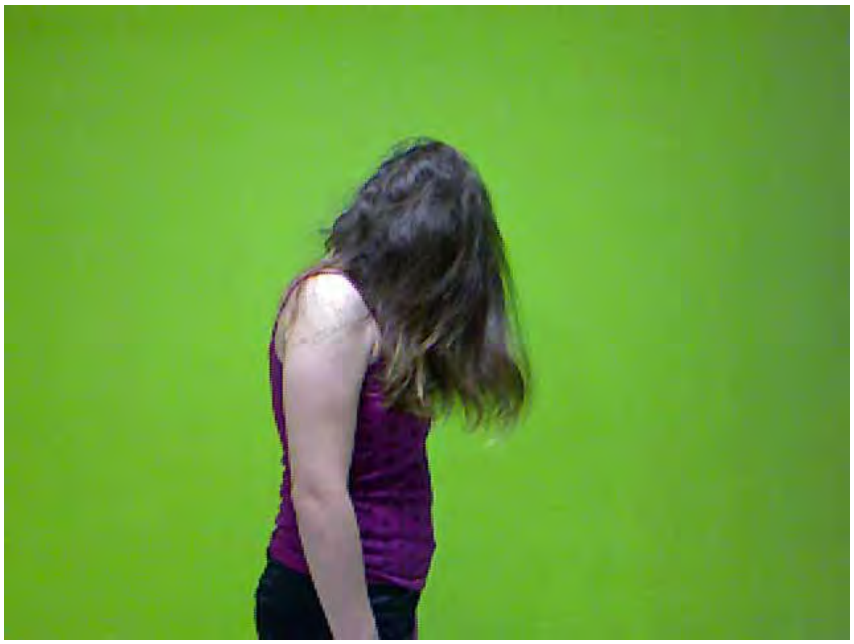


Figura A.37: *Img_5 - Imagem capturada pela câmera RGB*

Figura A.38: *Img_5 - Trimap manual*

Figura A.39: *Img_5 - Trimap automático*

Figura A.40: *Img_5 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.41: *Img_5 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.42: *Img_5 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.43: *Img_5 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.44: *Img_5 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.45: *Img_5 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

Figura A.46: *Img_6 - Imagem capturada pela câmera RGB*

Figura A.47: *Img_6 - Trimap manual*

Figura A.48: *Img_6 - Trimap automático*

Figura A.49: *Img_6 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.50: *Img_6 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.51: *Img_6 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.52: *Img_6 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.53: *Img_6 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.54: *Img_6 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

Figura A.55: *Img_7 - Imagem capturada pela câmera RGB*

Figura A.56: *Img_7 - Trimap manual*

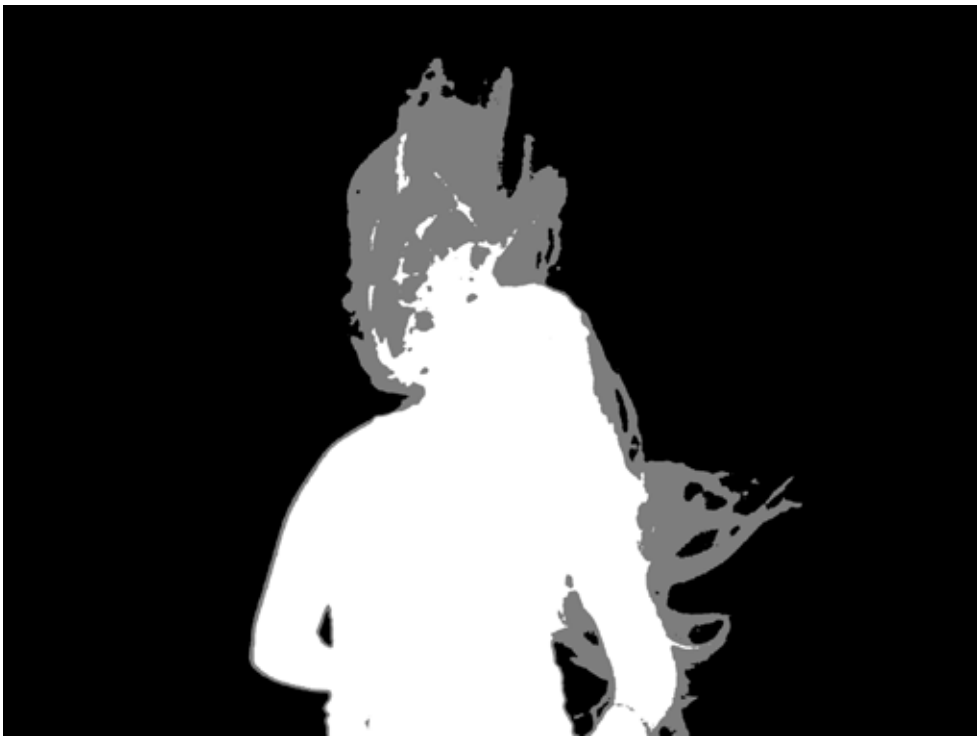


Figura A.57: *Img_7 - Trimap automático*

Figura A.58: *Img_7 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.59: *Img_7 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.60: *Img_7 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.61: *Img_7 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.62: *Img_7 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.63: *Img_7 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

Figura A.64: *Img_8 - Imagem capturada pela câmera RGB*

Figura A.65: *Img_8 - Trimap manual*

Figura A.66: *Img_8 - Trimap automático*

Figura A.67: *Img_8 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.68: *Img_8 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.69: *Img_8 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.70: *Img_8 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.71: *Img_8 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.72: *Img_8 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*

Figura A.73: *Img_9 - Imagem capturada pela câmera RGB*

Figura A.74: *Img_9 - Trimap manual*

Figura A.75: *Img_9 - Trimap automático*

Figura A.76: *Img_9 - Composição obtida com o Bayesian Matting a partir do Trimap manual*

Figura A.77: *Img_9 - Composição obtida com o Bayesian Matting a partir do Trimap automático*

Figura A.78: *Img_9 - Composição obtida com o Robust Matting a partir do Trimap manual*

Figura A.79: *Img_9 - Composição obtida com o Robust Matting a partir do Trimap automático*

Figura A.80: *Img_9 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap manual*

Figura A.81: *Img_9 - Composição obtida com o Geodesic Modificado Matting a partir do Trimap automático*